

# SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
 Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
 Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

## STAFF USE ONLY

Searcher: \_\_\_\_\_

Searcher Phone #: \_\_\_\_\_

Searcher Location: \_\_\_\_\_

Date Searcher Picked Up: 5/8/01

Date Completed: 5/8/01

Searcher Prep & Review Time: \_\_\_\_\_

Clerical Prep Time: \_\_\_\_\_

Online Time: \_\_\_\_\_

## Type of Search

NA Sequence (#) \_\_\_\_\_

AA Sequence (#) 1

Structure (#) \_\_\_\_\_

Bibliographic \_\_\_\_\_

Litigation \_\_\_\_\_

Fulltext \_\_\_\_\_

Patent Family \_\_\_\_\_

Other \_\_\_\_\_

## Vendors and cost where applicable

STN \_\_\_\_\_

Dialog \_\_\_\_\_

Questel/Orbit \_\_\_\_\_

Dr.Link \_\_\_\_\_

Lexis/Nexis \_\_\_\_\_

Sequence Systems 02

WWW/Internet \_\_\_\_\_

Other (specify) \_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2001, 11:54:40 ; Search time 12.28 Seconds  
(without alignments)  
769.686 Million cell updates/sec

Title: US-09-323-597b-2  
Perfect score: 2717  
Sequence: 1 MALNSGSPPAIGPYENHGX.....VYGVNVTWTDWYRQNRADG 492

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued\_Patents\_Aa.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfillseq1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2717	100.0	492	US-09-342-749-2	Sequence 2, Appl 1
2	1540	56.7	283	US-08-807-151-1	Sequence 1, Appl 1
3	676.5	24.9	435	US-09-008-271A-6	Sequence 6, Appl 1
4	660	24.3	798	US-08-200-900A-2	Sequence 2, Appl 1
5	558.5	20.6	855	PCT-US94-00616-2	Sequence 2, Appl 1
6	556	20.5	638	US-09-027-337-2	Sequence 2, Appl 1
7	539.5	19.9	416	US-08-681-151-3	Sequence 3, Appl 1
8	334	19.7	356	US-09-000-846-2	Sequence 2, Appl 1
9	328.5	19.5	256	US-08-681-151-1	Sequence 1, Appl 1
10	528.5	19.4	418	US-08-508-448C-25	Sequence 3, Appl 1
11	527.5	19.2	435	US-08-807-151-3	Sequence 25, Appl 1
12	521.5	19.1	812	US-08-248-629A-1	Sequence 3, Appl 1
13	518.5	19.1	812	US-08-451-932-1	Sequence 1, Appl 1
14	518.5	19.1	812	US-08-452-260-1	Sequence 1, Appl 1
15	518.5	19.1	812	US-08-326-785-1	Sequence 1, Appl 1
16	518.5	19.1	812	US-08-612-788-1	Sequence 1, Appl 1
17	518.5	19.1	812	US-08-605-598B-1	Sequence 1, Appl 1
18	518.5	19.1	812	US-08-429-743-1	Sequence 1, Appl 1
19	518.5	19.1	812	US-08-866-735-1	Sequence 1, Appl 1
20	518.5	19.1	812	US-09-066-028-1	Sequence 1, Appl 1
21	518.5	19.1	812	PCT-US95-05107-1	Sequence 1, Appl 1
22	513.5	18.9	232	US-08-508-448C-19	Sequence 19, Appl 1
23	513.5	16.9	314	US-09-008-271A-3	Sequence 3, Appl 1
24	459.5	16.6	546	5200340-6	Patent No. 5200340
25	452	16.6	270	US-08-978-404B-8	Sequence 8, Appl 1
26	450	16.6	635	US-08-148-910-12	Sequence 12, Appl 1
27	449	16.5	270	US-08-148-910-12	Sequence 12, Appl 1

28	449	16.5	655	US-08-448-937A-12	Sequence 12, Appl 1
29	447	16.5	276	US-09-016-366A-15	Sequence 15, Appl 1
30	447	16.5	276	US-08-978-404B-21	Sequence 21, Appl 1
31	446	16.4	230	US-08-379-621-2	Sequence 2, Appl 1
32	446	16.4	230	US-08-147-000B-2	Sequence 2, Appl 1
33	446	16.4	230	US-08-889-078-2	Sequence 2, Appl 1
34	445.5	16.4	791	US-08-643-219-1	Sequence 1, Appl 1
35	445.5	16.4	791	US-09-131-995-1	Sequence 1, Appl 1
36	445.5	16.4	791	US-08-832-087B-1	Sequence 1, Appl 1
37	445.5	16.4	791	US-08-831-350-1	Sequence 1, Appl 1
38	445.5	16.4	810	US-07-854-603-2	Sequence 2, Appl 1
39	445.5	16.4	810	US-08-147-000B-29	Sequence 29, Appl 1
40	445.5	16.4	810	5200340-8	Patent No. 5200340
41	444.5	16.4	790	US-08-469-486-54	Sequence 54, Appl 1
42	444.5	16.4	790	US-08-469-486-54	Sequence 54, Appl 1
43	442.5	16.3	222	US-08-456-840-46	Sequence 46, Appl 1
44	442.5	16.3	222	US-08-266-407A-46	Sequence 46, Appl 1
45	442.5	16.3	222	US-08-892-544-46	Sequence 46, Appl 1

## ALIGNMENTS

RESULT 1  
US-09-342-749-2  
; Sequence 2, Application US/09342749  
; Patent No. 6166194  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Alexander K.C.  
; APPLICANT: Tavligian, Sean V.  
; APPLICANT: Teng, David H.-F.  
; APPLICANT: Myriad Genetics, Inc.  
; TITLE OF INVENTION: TMPRSS2 is a Tumor Suppressor  
; FILE REFERENCE: 2318-202  
; CURRENT FILING DATE: 1999-06-29  
; EARLIER APPLICATION NUMBER: US 60/091,044  
; EARLIER FILING DATE: 1998-06-29  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patent Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-342-749-2

Query Match 100.0%; Score 2717; DB 4; Length 492;  
Best Local Similarity 100.0%; Pred. No. 2, 2e-241;  
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MALNSGSPPALGPYENHGXQREPNRYPAQPTVYVYVHQAQYPSAPVQYARVLTQA	60
Db	1	MALNSGSPPALGPYENHGXQREPNRYPAQPTVYVYVHQAQYPSAPVQYARVLTQA	60
QY	61	SNPVCTOPKPSSTGYCTSKTKKALCTLTGTFVGAALAAAGLTFKFMGSKCNSGIEC	120
Db	61	SNPVCTOPKPSSTGYCTSKTKKALCTLTGTFVGAALAAAGLTFKFMGSKCNSGIEC	120
QY	121	DSGTCINPNSMCGDVSHCHGEGEDENRCVRLYGNFTLOYVSSQKSMHVCODDNYET	180
Db	121	DSGTCINPNSMCGDVSHCHGEGEDENRCVRLYGNFTLOYVSSQKSMHVCODDNYET	180
QY	181	GRACRDMGYKNNFYSSOGIYDSDGSTSEFKLNTSAGNDVIYKKLHSDACSSKAVVSLR	240
Db	181	GRACRDMGYKNNFYSSOGIYDSDGSTSEFKLNTSAGNDVIYKKLHSDACSSKAVVSLR	240
QY	241	CIACGVNLSSROSRITYGGSALPGAMPQVSLHYOVHVGGSITPEVITAAHCVK	300
Db	241	CIACGVNLSSROSRITYGGSALPGAMPQVSLHYOVHVGGSITPEVITAAHCVK	300
QY	301	PLNNPMTAFAGILNOSFMFYAGYOVERVISHPNYDSKTKNNDLALMKQPLTFNDL	360
Db	301	PLNNPMTAFAGILNOSFMFYAGYOVERVISHPNYDSKTKNNDLALMKQPLTFNDL	360

Db 301 PLNPNHMTAFAGILRQSPMFYAGAYEVESHRYDSKTKNDIALMKLQKPLTFNDL 360  
Qy 361 VPRVCLPNEGMLQPEQLCWSMGATEEKGKTSYVLAIAKVLLETFQRCNSRYVDNL 420  
Db 361 VPRVCLPNEGMLQPEQLCWSMGATEEKGKTSYVLAIAKVLLETFQRCNSRYVDNL 420  
Qy 421 TRAMICAGLQGVNDSGCGDGGPLVTSKNNTWMLIGDTSWGSCKAKARPGYGVNMF 480  
Db 421 TRAMICAGLQGVNDSGCGDGGPLVTSKNNTWMLIGDTSWGSCKAKARPGYGVNMF 480  
Qy 481 TDWYRQMRADG 492  
Db 481 TDWYRQMRADG 492  
  
RESULT 2  
US-08-807-151-1  
; Sequence 1, Application US/08807151  
; Patent No. 6043033  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Lal, Preeti  
; TITLE OF INVENTION: NOVEL HUMAN PROSTATE-ASSOCIATED  
; TITLE OF INVENTION: PROTEASE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/807,151  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0227 US  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 283 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: SCORNOT01  
; CLONE: 556016  
; US-08-807-151-1  
  
Query Match 56.7%; Score 1540; DB 3; Length 283;  
Best Local Similarity 99.6%; Pred. No. 1.3e-133;  
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 61 QVSLHGVNHYVGGSLITPEWITVAHCEKPLNNPMHTAFAGILRQSPMFYAGAYE 120  
Qy 330 KYISHENYDSKTKNDIALMKLQKPLTFNDLVKPVCLPNEGMLQPEQLCWSMGATEE 389  
Db 121 KYISHENYDSKTKNDIALMKLQKPLTFNDLVKPVCLPNEGMLQPEQLCWSMGATEE 180  
Qy 390 KKTSEVLAIAKVLLETFQRCNSRYVDNLITPAMICAGLQGVNDSGCGDGGPLVTSK 449  
Db 181 KKTSEVLAIAKVLLETFQRCNSRYVDNLITPAMICAGLQGVNDSGCGDGGALVTSK 240  
Qy 450 NNIMWMLIGDTSWGSCKAKARPGYGVNMFYTDWYRQMRADG 492  
Db 241 NNIMWMLIGDTSWGSCKAKARPGYGVNMFYTDWYRQMRADG 283  
  
RESULT 3  
US-09-008-271A-6  
; Sequence 6, Application US/09008271A  
; Patent No. 6203979  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Yue, Henry  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Tang, Tom Y.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/008,271A  
; FILING DATE: 16-Jan-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: <Unknown>  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mohan-Peterson, Sheila  
; REGISTRATION NUMBER: 41,201  
; REFERENCE/DOCKET NUMBER: PF-0458 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 435 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: COLNOT13  
; CLONE: 1337018  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6 :  
US-09-008-271A-6  
  
Query Match 24.9%; Score 676.5; DB 4; Length 435;  
Best Local Similarity 39.1%; Pred. No. 4.8e-54;  
Matches 150; Conservative 57; Mismatches 128; Indels 49; Gaps 13;

Db 75 CDGELDPLGDEEHCVKSPFEGPAVAVRLSKDRSTLQVLDATGCMWESACDFNTEALA 134  
QY 182 RAACRDGY--KNNFYSQ-----GIYDSDGSTSPFKLNTSAGNVDIYKLYHSDACS 232  
Db 135 ETACRQMGYSKPFRAVEIGPDLDLVETENSQELRANSSG-----PCL 182  
QY 233 SKAVYSLRCLACGVNLNSSROSRIYVGSALPGAMPQVSLHYOVNHYVCGSIIITPENIV 292  
Db 183 SGLSVSLHCLACGSL--KTPRVYVGEASVDSMPQVSLQYDKQHCYCGSILDPHMYL 239  
QY 293 TAACVCEKPLNPNHMTAFAGILR-QSFMEFYAGYQVEKVI---SHRYPDKTKRNDAL 348  
Db 240 TAACHCFKRH-TDVNRMKVRAGSDKLSF---PSLAVALKIIIEFNPMY---PKNDIAL 291  
QY 349 MKLOKPLTFENDLVKPCLPNPGMMLQPEOLCWSMGATEEK-GKTSVYLNAAKYLLET 407  
Db 292 MKLOPLTFSTVAPICLPFDEDELTPATPLMTIGMGTCKNGMSIILQASQVIDS 351  
QY 408 QRCNSRYVDNLITPAMICAGFLQGNVDSGCGDGGPLVTSKNNIMLIGDTSMGSGCAK 467  
Db 352 TRCADADAYGEVEYKMKACGIPGEGVDTGCGDSGGLMYQSDQ-WHYVGIYSMGYCGG 410  
QY 468 AYRPGYGNVAFWDWIYKRMAD 491  
Db 411 PSTPGVYTKVSAYLNMWYINWKA 434

RESULT 4  
US-08-200-900A-2  
Sequence 2, Application US/08200900A  
Patent No. 5665566  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Genetics Institute, Inc. - Legal Affairs  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/200,900A  
FILING DATE: 23-FEB-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Melner, Maureen C.  
REGISTRATION NUMBER: 31,544  
REFERENCE/DOCKET NUMBER: GI 5201-FWC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 876-1170, X8574  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 798 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-200-900A-2

Query Match 24.3% Score 660; DB 1: Length 798;  
Best Local Similarity 33.9% Pred. No. 3.8e-52;  
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;  
QY 65 VCTQPSGPGVCTSKTKALCITLTL-----GTFLVGALAGLMLKFMGSKCSNSG 117

Db 374 VYTGPGPVNDVFSTTNRMVLEITDNMLAKGFRANFTTGLG-----IPECKEDN 426  
QY 118 IECDSGTCINPSNMCDGVSHCPGEGEDENRCVRLYG-----PNFLLQYSSQKRSWH 169  
Db 427 FQC-KDGEICILVNLCDGFPKCKDSDEAHCVRLFNGTTDSSGLVQFRIO-----SIWH 479  
QY 170 PVEDDMNENTGRAACRDGMYKNNFYSQGIYDSDGSTSPFKLNTSAGNVDIYKLYHSD 229  
Db 480 VACAENMTTQISDVCQLLGLGTG--NSSVPTFTSGGPPYVNLMTAPNGSLI---LTPSQ 534  
QY 220 ACSSKAVYSLRC--TAGVNLNLSNOS-RIVGGSALPGAMPQVSLHYOVNHYVCGSII 286  
Db 535 QLEDLSLLQCNKRSCKCKLVTQEVSPKIVGSGSREGAMPVVALFEDQVCGASLV 594  
QY 287 TPENITAAHCVEKPLNPNHMTAFAGILRQSFMEFYAGYQVE-----KVISHPYDSKT 341  
Db 535 SBDLVSAHACYGGRMPKSKKAVLGLHMASNL---TSPQETRLIQIYINPHYKRR 651  
QY 342 KNNDIALKLOKPLTFENDLVKPCLPNPGMMLQPEOLCWSMGATEEKGTSEVLNAK 401  
Db 652 KNNDIAMHLEKKNVYTDYIDPICLPEENQVFPGRICISAGMGALIVQGSTADVLQEAD 711  
QY 402 VLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSGCGDGGPLVTSKNNIMLIGDTSM 461  
Db 712 VPLLSNEKCOOQMPYXN-ITENMVCAGYEAGVDSGCGDGGPLMCOENRMMLLAGVTSF 770  
QY 462 GSGCAKAYRPGYGNVAFWDWI 484  
Db 771 GYOCALPNRPGVYARVPRFTWT 793

RESULT 5  
PCT-US94-00616-2  
Sequence 2, Application PC/TUS9400616  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE  
NUMBER OF SEQUENCES: 33  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/00616  
FILING DATE:  
CLASSIFICATION:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 798 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-00616-2

Query Match 24.3% Score 660; DB 5: Length 798;  
Best Local Similarity 33.9% Pred. No. 3.8e-52;  
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;  
QY 65 VCTQPSGPGVCTSKTKALCITLTL-----GTFLVGALAGLMLKFMGSKCSNSG 117  
Db 374 VYTGPGPVNDVFSTTNRMVLEITDNMLAKGFRANFTTGLG-----IPECKEDN 426  
QY 118 IECDSGTCINPSNMCDGVSHCPGEGEDENRCVRLYG-----PNFLLQYSSQKRSWH 169  
Db 427 FQC-KDGEICILVNLCDGFPKCKDSDEAHCVRLFNGTTDSSGLVQFRIO-----SIWH 479  
QY 170 PVEDDMNENTGRAACRDGMYKNNFYSQGIYDSDGSTSPFKLNTSAGNVDIYKLYHSD 229  
Db 480 VACAENMTTQISDVCQLLGLGTG--NSSVPTFTSGGPPYVNLMTAPNGSLI---LTPSQ 534



QY 351 LKRLPRLNDLVKPCVLPDNPGLMLOPEOLCHISGKATEEKGKTSSEVLNAKVLLETRC 410  
DB 489 LQPLPLNTEOKPLCPKSKADNTITNCWYTGWYKKEGETONILQKATPLVPEEC 548  
QY 411 NSRYVDNLTPAMICAGLOGVNDSCGSGPLVSKNNIMWLGDTSGSCAKAYR 470  
DB 549 QKRT-KRYVITKQKICGICGEGGIDACKDGSGLVCKHSGRMOLVITTSWEGCAREQ 607  
QY 471 PGVYGNVMTFTDWTYRORA 490  
DB 608 PGVYTKVAEYIDWLEKIOS 627

RESULT 8  
US-09-000-846-2  
Sequence 2, Application US/09000846  
Patent No. 5981830  
GENERAL INFORMATION:  
APPLICANT: WU, QINGYU  
APPLICANT: SADLER, JASPER  
TITLE OF INVENTION: KNOCKOUT MICE AND THEIR PROGENY WITH  
TITLE OF INVENTION: DISRUPTED SERINE PROTEASE GENES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.  
STREET: 2200 CLARENDON BLVD. SUITE 1400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: US  
ZIP: 22201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/000,846  
FILING DATE: 30-DEC-1997  
CLASSIFICATION: 800  
Prior Application DATA:  
APPLICATION NUMBER: US/08/866,058  
FILING DATE: 30-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: LEBOVITZ, RICHARD M.  
REGISTRATION NUMBER: 37,067  
REFERENCE/DOCKET NUMBER: BERLX 65P1  
TELEPHONE: 703-243-6333  
TELEFAX: 703-243-6410  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 416 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-000-846-2

Query Match 19.98; Score 539.5; DB 2; Length 416;  
Best Local Similarity 30.68; Pred. No. 1.7e-41;  
Matches 132; Conservative 64; Mismatches 179; Indels 57; Gaps 9;

QY 77 CTSKTKKALCITLTGLFELVAGALAGLMLKFKGSKNSGIEDSSGTCINSMWCDGV 136  
DB 11 CCSRPVVALIVGT-LFLTG-----IGASNAIVT 40  
QY 137 SHCPGGEDENCVRLYGNPILQVYSQKSMHPVCCDDNNENYGRAACRDMGKNNFYS 196  
DB 41 ILQSQDEPLVQVLSGDRSLAVFDTEGTWRLCSSRNARVAGCGEMGLRALAH 100  
QY 197 SGCIVDDSGSTSMKLTSA-----GNVDIYKKLYHSDA-----CSSAAVYSLRCIACGVN 247  
DB 101 SELDVRTAGAN-----GTSGFVCEVDEGLRLAORLLDIVISVCDPGRGFLTATCODCG-- 153

QY 248 LNSSROSRIYGESALPGAMPQVSLHVQNVHCGSITPEIVTAACVERPLNPMH 307  
DB 154 RRLPLVRIYVIGODSSLGRRPMQVSLRYDGTHLGGSLLSGDWLTPAARCFPERNNLSR 213  
QY 308 WTAFAGLTRSEFMFYAGYQYEVKYSHPN-----DSKKNNDIAMLKQKRLTNDLV 361  
DB 214 WRFVAGAVARTSP-HAVOLVQAVIYHGTLPPRDPPTIDENSDIALVHLSSSLPLEYI 272  
QY 362 KPVCLPNDPMMLOPEOLCMTISGKATEEKGKTSSEVLNAKVLLETRCNSRYVDNLIT 421  
DB 273 QPVCLPAAQALVDGKCYVTGNGNTQFYGQAMVLOEARVPIISNVCNSPDTGNOIK 332  
QY 422 PAMICAGLOGVNDSCGSGGPLYV---TSKNNIMWLGDTSGSCAKAYRPGVYGNV 477  
DB 333 PKMECAGYPEGGIDACGDSGGPFVCEDSISGSRMRLCIGVSMGTCALARRPGVYTKV 392  
QY 478 WFTDWTYRORA 490  
DB 393 TDFREWIFKAIR 404

RESULT 9  
US-08-681-151-1  
Sequence 1, Application US/08681151  
Patent No. 5869637  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Bandman, Olga  
APPLICANT: Braxton, Scott Michael  
TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/681,151  
FILING DATE: Herewith  
CLASSIFICATION: 435  
Prior Application DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0074US  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 356 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: HEARNOT01  
CLONE: 307474  
US-08-681-151-1

RA Oberst M.D., Dickson R.B., Lin C.Y.  
 RN Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RL [1]  
 RP SEQUENCE FROM N.A.  
 RA Takeuchi T., Shuman M.A., Craik C.S.;  
 RT "Reverse biochemistry: Use of macromolecular protease inhibitors to  
 RT dissect complex biological processes and identify a membrane-type  
 RT serine protease in epithelial cancer and normal tissue."  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).  
 DR EMBL: AF118224; AAD2765.2; -  
 DR EMBL: AF133086; AAF00109.1; -  
 DR HSP: P00763; IDPO.  
 DR MEROPS: S01.302; -  
 DR INTERPRO: IPRO00859; -  
 DR INTERPRO: IPRO01254; -  
 DR INTERPRO: IPRO01314; -  
 DR INTERPRO: IPRO02172; -  
 DR PFM: PF00057; Idl\_recept\_a; 4.  
 DR PFM: PF00089; trypsin; 1.  
 DR PFM: PF00431; CUB; 2.  
 DR PRINTS: PR00261; LDLRECEPTOR.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 DR PROSITE: PS01209; LDLR\_1; 2.  
 DR PROSITE: PS50068; LDLR\_2; 4.  
 KM Glycoprotein; Protease.  
 SO SEQUENCE 855 AA; 94769 MW; 2614132C01F9C9 CRC64;

Query Match 20.6%; Score 558.5; DB 4; Length 855;  
 Best Local Similarity 33.9%; Pred. No. 3.6e-42;  
 Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;

OY 110 GSKNSGIECDSCGTCINPSCMGDGVSHCGEGDENRCVLTGYPNFIQVYSSQKSMH 169  
 Db 522 GCSPAPATFRC-SNGKLSKSGQCGKDDCGSDDEASCPV---NVY----- 565  
 OY 170 PVCDDMNENGRACRDMGKNNFYSSQGIYDSSGTSFKMLNTSAGNDVIKLYHSD 229  
 Db 566 -----TCT-----KHTYRCLNG-----LCLSKGNPDCGKEDCSD 595  
 OY 230 ACSSKAYVSLACICGVVNS-SQSRVYGESALPGAMPQVSLH-VQNVHVGGSIT 287  
 Db 596 GSDEK-----CCDCGLRSFTROARVGGTADDEGMPQVSLHAGOHIGCASLIS 647  
 OY 288 PEMVTAHACVEP-----LNPVMTAFAGILRQSFMYGAGY---VEKYISHPNYSK 340  
 Db 648 PNLVSAHCHTIDRGFRYSPTQMTAFGLHDS-QRSAPGVQERRIKRLIHPFFWDF 706  
 OY 341 TKNDIALMLKLOKPLTFNDLVKPYCLPMPGMLPEOLCMISSNGATEEKGKTSFVLNA 400  
 Db 707 TFDVDIALLEEKAEVSSVWRPICLPDASHVFPAGKAIWVTGNGHQYGTGALLDKG 766  
 OY 401 KVLLEIQRCSRYVYDNLITPAMICAGFLGAGNDSCGSDGSGPL--VTSNNNTWMLGD 458  
 Db 767 EIRVINTOTEN--LTPQITPRMCAVGLSGVSCGSDGGLSSVEADGRI-FQAGV 823  
 OY 459 TSWGCAKAYRPGVYGVVWVETDMI 484  
 Db 824 VSWGDGCAQRKPKPVYTRLPFRDHI 849

RESULT 11

OYVANG PRELIMINARY; PRT; 1379 AA.

AC 03VANG;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE CG2105 PROTEIN.  
 GN Drosophila melanogaster (Fruit fly).

CC Eukaryota: Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 ON NCBI\_Taxid=7227;

RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; Pubmed=10731132;  
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Baellon R.V., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Doolin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,  
 RA Foutel C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gottlieb J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshire A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AEO03840; AAF59230.1; -  
 DR HSP: P00763; IDPO.  
 DR FLTBASE: FBgn0033192; CG2105.  
 DR INTERPRO: IPRO00024; -  
 DR INTERPRO: IPRO00024; -  
 DR INTERPRO: IPRO00217; -  
 DR INTERPRO: IPRO01190; -  
 DR INTERPRO: IPRO01254; -  
 DR INTERPRO: IPRO01314; -  
 DR INTERPRO: IPRO02172; -  
 DR PFM: PF00057; Idl\_recept\_a; 2.  
 DR PFM: PF00089; trypsin; 1.  
 DR PFM: PF01392; Fz; 1.  
 DR PRINTS: PR00261; LDLRECEPTOR.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 DR PROSITE: PS00227; TUBULIN; 1.  
 DR PROSITE: PS01209; LDLR\_1; 1.  
 DR PROSITE: PS50068; LDLR\_2; 2.  
 SO SEQUENCE 1379 AA; 149489 MW; A593A9CC2167E4B7 CRC64;

Query Match 20.0%; Score 544; DB 5; Length 1379;

Best Local Similarity 33.4%; Pred. No. 1.4e-40;  
 Matches 141; Conservative 65; Mismatches 158; Indels 58; Gaps 20;

OY 111 SKNSNG---IECDSCGTCINPSCMGDGVSHCGEGDENRCVLTGYPNFIQVYSSQKSMH 162











```
Db 699 TICYITGMEGTGTFGAGRLKE---AQLPVIENKVCNRVEYLNNRKYKSTELCAGOLAGG 754
QY 434 VDSCGDSGGPLVTSKNINIMLIGDTSMSGCAKAYRPGYGNMVFETDNIYRQMR 489
Db 755 VDSCGDSGGPLVCFERDKYILOGVTSMLGCAKPNKPGYVVRYSREVDWIEREMR 810
```

Search completed: May 8, 2001, 11:55:00  
Job time: 20 sec



Query Match	Best Local Similarity	Score	DB 1	Length	638
Matches 115	Conservative 52	Mismatches 117	Indels 26	Gaps	
QY 185 CRDMGYKNNFYSSGICVDDSGSTFNMKLTSAANDIYKKLYHSDACSSKAAVSLRCLAC	19.8%	37.1%	Pred. No. 8.3e-33		
DB 340 CKEECCKSLRLS---TDGSPRTITYGMGSSGYSRLCKLVDSPDCTTKI-----					387
QY 245 GVNLSNNSQSRIVYGESALPGAMPQVSLHYQNV---HYCGSIIIPRENIYAAHCEVKP					301
DB 388 -----NARIVGSTNASLGEPMQVSLQVLAVSOTHLGCGSIIIGROWVLAHCFD-G					438
QY 302 LNNPHTWAFAGILIQSEFMYGA-GYOVEKIVSHPYDSTKKNIDIALKLOKPLTFNDL					360
DB 439 IPRYVWMIITGGILSELSETRKTPSPSRITRELLIHQEKYKSEGNIDIALKLOKPLWYTFE					498
QY 361 VKPYCLPNPQMLLOPEOLCOWISGWAITEKGKTSSEVLANAKVLLIETORCNSRYVYDNL					420
DB 499 OKPICLPKADNTITYTNCWYTGWGYTRKQEGEITONILAKATIPLVNRECQKKY-RDYVI					557
QY 421 TPAMTCAEFLQGVNDSGCGDSCGPLVTSKNNIIMWLIGDTSMSGCKAKARPGVYGNVMP					480
DB 558 NKOMICAGYKEGTDACGKDGSGPLVCKHSGHGMOLVGLTSMGEGCGKRDGPVYTYVSEY					617
QY 481 TDWITROMRA 490					
DB 618 MDWILEKTOS 627					

```

01 RESULT 14
02 HEP5_MOUSE
03 ID HEP5_MOUSE STANDARD; PRT: 416 AA.
04 AC 035453;
05 DT 15-JUL-1998 (Rel. 36, Created)
06 DT 15-JUL-1998 (Rel. 36, Last sequence update)
07 DT 15-JUL-1998 (Rel. 36, Last annotation update)
08 DE SERINE PROTEASE HEP5IN (EC 3.4.21.-)
09 HPN.
10 OS Mus musculus (Mouse).
11 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
12 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
13 CC NCBI_TaxID=10090;
14 RN [1]
15 RP SEQUENCE FROM N.A.
16 RC TISSUE-Layer:
17 RA MEDLINE=98058912; PubMed=9395459;
18 RA Vu T.K.H., Liu R.W., Haaksma C., Tomasek J.J., Howard E.W.;
19 RT "Identification and cloning of the membrane-associated serine
20 RT protease, hepsin, from mouse preimplantation embryos.";
21 RL J. Biol. Chem. 272:31315-31320(1997).
22 CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
23 CC OF CELL MORPHOLOGY.
24 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
25 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
26 CC TRYPSIN FAMILY.
27 CC -----
28 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
29 CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
30 CC the European Bioinformatics Institute. There are no restrictions on its
31 CC use by non-profit institutions as long as its content is in no way
32 CC modified and this statement is not removed. Usage by and for commercial
33 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
34 CC or send an email to license@isb-sib.ch).
35 CC -----
36 CC EMBL: AF030065; AAB84221.1; -
37 CC HSSP: P00763; IDPO.
38 CC MEROPS: S01.224; -.
39 CC MGD: MGI:1196620; Hpn.
40 CC InterPro: IPR001254; -
41 CC InterPro: IPR001314; -
42 CC Pfam: PF000089; trypsin; 1.
43 CC PRINTS: PRO0722; CHYMOTRYPSIN.
44 CC PROSITE: PS00134; TRYPSIN_HIS. 1.
45 CC PROSITE: PS00135; TRYPSIN_SER. 1.
46 CC KW Hydroxylase; Serine protease; Transmembrane; Signal-anchor.
47 CC CHAIN 1 161
48 CC CHAIN 162 416
49 CC CHAIN 1 16
50 CC DOMAIN 1 16
51 CC TRANSMEM 17 43
52 FT FT
53 FT FT
54 FT FT
55 FT FT
56 FT FT
57 FT FT
58 FT FT
59 FT FT
60 FT FT
61 FT FT
62 FT FT
63 FT FT
64 FT FT
65 FT FT
66 FT FT
67 FT FT
68 FT FT
69 FT FT
70 FT FT
71 FT FT
72 FT FT
73 FT FT
74 FT FT
75 FT FT
76 FT FT
77 FT FT
78 FT FT
79 FT FT
80 FT FT
81 FT FT
82 FT FT
83 FT FT
84 FT FT
85 FT FT
86 FT FT
87 FT FT
88 FT FT
89 FT FT
90 FT FT
91 FT FT
92 FT FT
93 FT FT
94 FT FT
95 FT FT
96 FT FT
97 FT FT
98 FT FT
99 FT FT
100 FT FT
101 FT FT
102 FT FT
103 FT FT
104 FT FT
105 FT FT
106 FT FT
107 FT FT
108 FT FT
109 FT FT
110 FT FT
111 FT FT
112 FT FT
113 FT FT
114 FT FT
115 FT FT
116 FT FT
117 FT FT
118 FT FT
119 FT FT
120 FT FT
121 FT FT
122 FT FT
123 FT FT
124 FT FT
125 FT FT
126 FT FT
127 FT FT
128 FT FT
129 FT FT
130 FT FT
131 FT FT
132 FT FT
133 FT FT
134 FT FT
135 FT FT
136 FT FT
137 FT FT
138 FT FT
139 FT FT
140 FT FT
141 FT FT
142 FT FT
143 FT FT
144 FT FT
145 FT FT
146 FT FT
147 FT FT
148 FT FT
149 FT FT
150 FT FT
151 FT FT
152 FT FT
153 FT FT
154 FT FT
155 FT FT
156 FT FT
157 FT FT
158 FT FT
159 FT FT
160 FT FT
161 FT FT
162 FT FT
163 FT FT
164 FT FT
165 FT FT
166 FT FT
167 FT FT
168 FT FT
169 FT FT
170 FT FT
171 FT FT
172 FT FT
173 FT FT
174 FT FT
175 FT FT
176 FT FT
177 FT FT
178 FT FT
179 FT FT
180 FT FT
181 FT FT
182 FT FT
183 FT FT
184 FT FT
185 FT FT
186 FT FT
187 FT FT
188 FT FT
189 FT FT
190 FT FT
191 FT FT
192 FT FT
193 FT FT
194 FT FT
195 FT FT
196 FT FT
197 FT FT
198 FT FT
199 FT FT
200 FT FT
201 FT FT
202 FT FT
203 FT FT
204 FT FT
205 FT FT
206 FT FT
207 FT FT
208 FT FT
209 FT FT
210 FT FT
211 FT FT
212 FT FT
213 FT FT
214 FT FT
215 FT FT
216 FT FT
217 FT FT
218 FT FT
219 FT FT
220 FT FT
221 FT FT
222 FT FT
223 FT FT
224 FT FT
225 FT FT
226 FT FT
227 FT FT
228 FT FT
229 FT FT
230 FT FT
231 FT FT
232 FT FT
233 FT FT
234 FT FT
235 FT FT
236 FT FT
237 FT FT
238 FT FT
239 FT FT
240 FT FT
241 FT FT
242 FT FT
243 FT FT
244 FT FT
245 FT FT
246 FT FT
247 FT FT
248 FT FT
249 FT FT
250 FT FT
251 FT FT
252 FT FT
253 FT FT
254 FT FT
255 FT FT
256 FT FT
257 FT FT
258 FT FT
259 FT FT
260 FT FT
261 FT FT
262 FT FT
263 FT FT
264 FT FT
265 FT FT
266 FT FT
267 FT FT
268 FT FT
269 FT FT
270 FT FT
271 FT FT
272 FT FT
273 FT FT
274 FT FT
275 FT FT
276 FT FT
277 FT FT
278 FT FT
279 FT FT
280 FT FT
281 FT FT
282 FT FT
283 FT FT
284 FT FT
285 FT FT
286 FT FT
287 FT FT
288 FT FT
289 FT FT
290 FT FT
291 FT FT
292 FT FT
293 FT FT
294 FT FT
295 FT FT
296 FT FT
297 FT FT
298 FT FT
299 FT FT
300 FT FT
301 FT FT
302 FT FT
303 FT FT
304 FT FT
305 FT FT
306 FT FT
307 FT FT
308 FT FT
309 FT FT
310 FT FT
311 FT FT
312 FT FT
313 FT FT
314 FT FT
315 FT FT
316 FT FT
317 FT FT
318 FT FT
319 FT FT
320 FT FT
321 FT FT
322 FT FT
323 FT FT
324 FT FT
325 FT FT
326 FT FT
327 FT FT
328 FT FT
329 FT FT
330 FT FT
331 FT FT
332 FT FT
333 FT FT
334 FT FT
335 FT FT
336 FT FT
337 FT FT
338 FT FT
339 FT FT
340 FT FT
341 FT FT
342 FT FT
343 FT FT
344 FT FT
345 FT FT
346 FT FT
347 FT FT
348 FT FT
349 FT FT
350 FT FT
351 FT FT
352 FT FT
353 FT FT
354 FT FT
355 FT FT
356 FT FT
357 FT FT
358 FT FT
359 FT FT
360 FT FT
361 FT FT
362 FT FT
363 FT FT
364 FT FT
365 FT FT
366 FT FT
367 FT FT
368 FT FT
369 FT FT
370 FT FT
371 FT FT
372 FT FT
373 FT FT
374 FT FT
375 FT FT
376 FT FT
377 FT FT
378 FT FT
379 FT FT
380 FT FT
381 FT FT
382 FT FT
383 FT FT
384 FT FT
385 FT FT
386 FT FT
387 FT FT
388 FT FT
389 FT FT
390 FT FT
391 FT FT
392 FT FT
393 FT FT
394 FT FT
395 FT FT
396 FT FT
397 FT FT
398 FT FT
399 FT FT
400 FT FT
401 FT FT
402 FT FT
403 FT FT
404 FT FT
405 FT FT
406 FT FT
407 FT FT
408 FT FT
409 FT FT
410 FT FT
411 FT FT
412 FT FT
413 FT FT
414 FT FT
415 FT FT
416 FT FT
417 FT FT
418 FT FT
419 FT FT
420 FT FT
421 FT FT
422 FT FT
423 FT FT
424 FT FT
425 FT FT
426 FT FT
427 FT FT
428 FT FT
429 FT FT
430 FT FT
431 FT FT
432 FT FT
433 FT FT
434 FT FT
435 FT FT
436 FT FT
437 FT FT
438 FT FT
439 FT FT
440 FT FT
441 FT FT
442 FT FT
443 FT FT
444 FT FT
445 FT FT
446 FT FT
447 FT FT
448 FT FT
449 FT FT
450 FT FT
451 FT FT
452 FT FT
453 FT FT
454 FT FT
455 FT FT
456 FT FT
457 FT FT
458 FT FT
459 FT FT
460 FT FT
461 FT FT
462 FT FT
463 FT FT
464 FT FT
465 FT FT
466 FT FT
467 FT FT
468 FT FT
469 FT FT
470 FT FT
471 FT FT
472 FT FT
473 FT FT
474 FT FT
475 FT FT
476 FT FT
477 FT FT
478 FT FT
479 FT FT
480 FT FT
481 FT FT
482 FT FT
483 FT FT
484 FT FT
485 FT FT
486 FT FT
487 FT FT
488 FT FT
489 FT FT
490 FT FT
491 FT FT
492 FT FT
493 FT FT
494 FT FT
495 FT FT
496 FT FT
497 FT FT
498 FT FT
499 FT FT
500 FT FT
501 FT FT
502 FT FT
503 FT FT
504 FT FT
505 FT FT
506 FT FT
507 FT FT
508 FT FT
509 FT FT
510 FT FT
511 FT FT
512 FT FT
513 FT FT
514 FT FT
515 FT FT
516 FT FT
517 FT FT
518 FT FT
519 FT FT
520 FT FT
521 FT FT
522 FT FT
523 FT FT
524 FT FT
525 FT FT
526 FT FT
527 FT FT
528 FT FT
529 FT FT
530 FT FT
531 FT FT
532 FT FT
533 FT FT
534 FT FT
535 FT FT
536 FT FT
537 FT FT
538 FT FT
539 FT FT
540 FT FT
541 FT FT
542 FT FT
543 FT FT
544 FT FT
545 FT FT
546 FT FT
547 FT FT
548 FT FT
549 FT FT
550 FT FT
551 FT FT
552 FT FT
553 FT FT
554 FT FT
555 FT FT
556 FT FT
557 FT FT
558 FT FT
559 FT FT
560 FT FT
561 FT FT
562 FT FT
563 FT FT
564 FT FT
565 FT FT
566 FT FT
567 FT FT
568 FT FT
569 FT FT
570 FT FT
571 FT FT
572 FT FT
573 FT FT
574 FT FT
575 FT FT
576 FT FT
577 FT FT
578 FT FT
579 FT FT
580 FT FT
581 FT FT
582 FT FT
583 FT FT
584 FT FT
585 FT FT
586 FT FT

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: May 8, 2001, 11:54:40 ; Search time 23.01 seconds  
(without alignments)  
1222.265 Million cell updates/sec

Title: US-09-323-597B-2

Perfect score: 2717  
Sequence: 1 MALNSGSPPAIGPYENHGY.....VYGNVMTFTWYRQRADG 492

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /SID56/gcgdata/geneseq/geneseqp/AA1980.DAT:\*  
2: /SID56/gcgdata/geneseq/geneseqp/AA1981.DAT:\*  
3: /SID56/gcgdata/geneseq/geneseqp/AA1982.DAT:\*  
4: /SID56/gcgdata/geneseq/geneseqp/AA1983.DAT:\*  
5: /SID56/gcgdata/geneseq/geneseqp/AA1984.DAT:\*  
6: /SID56/gcgdata/geneseq/geneseqp/AA1985.DAT:\*  
7: /SID56/gcgdata/geneseq/geneseqp/AA1986.DAT:\*  
8: /SID56/gcgdata/geneseq/geneseqp/AA1987.DAT:\*  
9: /SID56/gcgdata/geneseq/geneseqp/AA1988.DAT:\*  
10: /SID56/gcgdata/geneseq/geneseqp/AA1989.DAT:\*  
11: /SID56/gcgdata/geneseq/geneseqp/AA1990.DAT:\*  
12: /SID56/gcgdata/geneseq/geneseqp/AA1991.DAT:\*  
13: /SID56/gcgdata/geneseq/geneseqp/AA1992.DAT:\*  
14: /SID56/gcgdata/geneseq/geneseqp/AA1993.DAT:\*  
15: /SID56/gcgdata/geneseq/geneseqp/AA1994.DAT:\*  
16: /SID56/gcgdata/geneseq/geneseqp/AA1995.DAT:\*  
17: /SID56/gcgdata/geneseq/geneseqp/AA1996.DAT:\*  
18: /SID56/gcgdata/geneseq/geneseqp/AA1997.DAT:\*  
19: /SID56/gcgdata/geneseq/geneseqp/AA1998.DAT:\*  
20: /SID56/gcgdata/geneseq/geneseqp/AA1999.DAT:\*  
21: /SID56/gcgdata/geneseq/geneseqp/AA2000.DAT:\*  
22: /SID56/gcgdata/geneseq/geneseqp/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	2717	100.0	492 21	Y77726 Human tumour suppressor
2	2717	100.0	492 21	Y44406 Human 20P1F12-GTC2
3	2709	99.7	492 21	Y92050 HTPCa6/7 polypeptide
4	2696	99.2	492 21	B36801 Human TMPS22 prot
5	2696	99.2	492 21	Y57280 OVR15 homolog pro
6	1540	56.7	283 21	Y81492 Human prostate-acc
7	888	32.7	454 21	B32246 Tumour associated
8	885	32.6	452 20	Y41694 Human PRO382 prote
9	884	32.5	453 21	B44250 Human PRO382 (UNQ3
10	742.5	27.3	327 21	Y72093 Human serine prote
11	676.5	24.9	435 20	Y06437 Human protease HUP

12	674	24.8	432	21	Y99417 Human PRO1570 (UNQ
13	666.5	24.5	480	21	B08950 Human secreted pro
14	665.5	24.5	445	21	B11698 Mouse serine prote
15	660	24.3	798	15	R57283 Bovine enterokinas
16	651.5	24.0	457	21	B11699 Human serine prote
17	649.5	23.9	248	21	B43572 Human cancer assoc
18	648.5	23.9	414	21	B08912 Human secreted pro
19	601	22.1	1042	21	Y44426 Human serine prote
20	588	21.6	311	21	B11696 Mouse serine prote
21	587.5	21.6	273	21	B11696 Mouse serine prote
22	577.5	21.3	1113	21	Y44427 Human matrilysin (
23	558.5	20.6	683	21	B19551 Human peptidase, H
24	558.5	20.6	762	21	Y90284 Human antigen der
25	558.5	20.6	855	20	Y06671 Human matrilysin
26	558.5	20.6	855	21	B19552 Human serine prote
27	558	20.5	296	21	Y72108 Human serine prote
28	558	20.5	372	21	Y72092 Mouse serine prote
29	544	20.0	238	21	B11695 Mouse hepsin prote
30	539.5	19.9	416	20	Y43325 A mouse serine pro
31	539.5	19.9	416	20	W96812 Human PRO618 prote
32	538.5	19.8	802	20	Y41710 Human PRO618 (UNQ3
33	538.5	19.8	802	21	B44266 Human PRO618 prote
34	538.5	19.8	802	21	B24052 Human PRO618 prote
35	534	19.7	356	19	W46917 Amino acid sequenc
36	527.5	19.4	418	17	R89435 Trypsin-like enzym
37	527.5	19.4	418	20	Y29498 Human lung tumour
38	527.5	19.4	418	20	Y29501 Human lung tumour
39	527.5	19.4	418	20	Y29502 Human lung tumour
40	527.5	19.4	418	21	B44437 Human lung tumour
41	527.5	19.4	418	21	B44438 Human lung tumour
42	526.5	19.4	418	21	B44428 Human lung tumour
43	518.5	19.1	812	16	R83959 Complete mouse pla
44	518.5	19.1	812	17	W07585 Murine plasminogen
45	518.5	19.1	812	20	Y08686 Murine plasminogen

#### ALIGNMENTS

RESULT	1	Y77726	standard; Protein; 492 AA.
ID	Y77726		
XX	Y77726		
AC	Y77726		
XX	12-MAY-2000 (first entry)		
DT	12-MAY-2000 (first entry)		
XX	Human tumour suppressor TMPS22 polypeptide.		
DE	Human tumour suppressor gene; TMPS22; cancer; drug design;		
XX	Tumour suppressor gene; TMPS22; cancer; drug design;		
KW	gene therapy; protein therapy.		
KM	gene therapy; protein therapy.		
XX	Homo sapiens.		
OS	Homo sapiens.		
XX	WO20000605-A1.		
FN	WO20000605-A1.		
XX	06-JAN-2000.		
PD	06-JAN-2000.		
XX	29-JUN-1999; 99WO-US14622.		
PF	29-JUN-1999; 99WO-US14622.		
XX	29-JUN-1998; 98US-0091044.		
PR	29-JUN-1998; 98US-0091044.		
XX	(MYRI-) MYRIAD GENETICS INC.		
PA	(MYRI-) MYRIAD GENETICS INC.		
XX	Wong AKC, Tavtigian SV, Teng DHF;		
PI	Wong AKC, Tavtigian SV, Teng DHF;		
XX	WPI: 2000-170914/15.		
DR	WPI: 2000-170914/15.		
XX	N-PSDB: 287786.		
XX	Novel tumor suppressor TMPS22 used for the diagnosis and prognosis of		
PT	human cancer -		
XX	human cancer -		
PS	Claim 55; Page 77-79; 89pp; English.		





Db 348 FLRLSMGSPTR---AYGTGSSG---YSLRLCMTGDNVCTTKSTRIVGNTSSMGE 401  
 QY 267 WPMQVSLHVO---NHVCGSSITTPETIVTAHCV- KPLNPHMTAFAGILROSMFY 322  
 Db 402 WPMQVSLHVO---NHVCGSSITTPETIVTAHCV- KPLNPHMTAFAGILROSMFY 322  
 QY 323 GAGY-QVEKISHPNYDSKKNNDIALMKLOKPLTFENDLKPVCPLPBGMLDPEQCLWI 381  
 Db 460 DTFPSQIKETIIHQNKVNSGNHDIKLOAPLNTYEFQKPCICLPKGTSTIYTCMV 519  
 QY 382 SCGWATEEKSEKTEVLNAKAVLLIETORCNSRYVDNLITPAMICAGFLQGVDSOCGDS 441  
 Db 520 TGMGFSKEKEIQLNLOKVPILVTNECCGRY-QDKITQRYKACGKRGADACGDS 578  
 QY 442 GGPLVYSKNNIMWLGDTSMGSCAKAYRPGVYGVNVTFTDIYRMR-ADG 492  
 Db 579 GGPLVYSKNNIMWLGDTSMGSCAKAYRPGVYGVNVTFTDIYRMR-ADG 492  
 RESULT 9  
 HEPES\_HUMAN STANDARD; PRT; 417 AA.  
 AC POS981;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE SERINE PROTEASE HEPESIN (EC 3.4.21.-) (TRANSMEMBRANE PROTEASE, SERINE  
 1).  
 DE HPN OR TMPRSS1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA MEDLINE=88309431; PubMed=2835076;  
 RA Leytus S.P., Loeb K.R., Hagen F.S., Kurachi K., Davle E.W.;  
 RA "A novel trypsin-like serine protease (hepsin) with a putative  
 RA transmembrane domain expressed by human liver and hepatoma cells.";  
 RA Biochemistry 27:1067-1074(1988).  
 RL (2)  
 RP CHARACTERIZATION.  
 RA MEDLINE=91358502; PubMed=1885621;  
 RA Tsui J.A., Torres-Rosado A., Arai T., le Beau M.M., Lemons R.S.,  
 RA Chou S.H., Kurachi K.;  
 RA "Hepsin, a cell membrane-associated protease. Characterization,  
 RA tissue distribution, and gene localization";  
 RA J. Biol. Chem. 266:16948-16953(1991).  
 RL (3)  
 RP CHARACTERIZATION.  
 RA MEDLINE=93348237; PubMed=8346233;  
 RA Torres-Rosado A., O'Shea K.S., Tsui J.A., Chou S.H., Kurachi K.;  
 RA "Hepsin, a putative cell-surface serine protease, is required for  
 RA mammalian cell growth.";  
 RA Proc. Natl. Acad. Sci. U.S.A. 90:7181-7187(1993).  
 CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE  
 CC OF CELL MORPHOLOGY.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: PRESENT IN MOST TISSUES, WITH THE HIGHEST  
 CC LEVEL IN LIVER.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 DR EMBL, M18930; AAA36013.1; -

DR EMBL: X07732; CAA30558.1; -  
 DR EMBL: X07002; CAA30058.1; -  
 DR EMBL: S00845; S00845.  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.224; -  
 DR MIM: 142440; -  
 DR InterPro: IPR001254; -  
 DR InterPro: IPR001314; -  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PF00722; CHYMOTRYPSIN.  
 DR PROSITE: PS00134; TRYPSIN\_HIS.1.  
 DR PROSITE: PS00135; TRYPSIN\_SER.1.  
 KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor.  
 FT CHAIN 1 162  
 FT CHAIN 163 417  
 FT DOMAIN 1 17  
 FT TRANSMEM 18 44  
 FT  
 FT DOMAIN 45 417  
 FT DOMAIN 163 417  
 FT ACT\_SITE 203 257  
 FT ACT\_SITE 257 257  
 FT ACT\_SITE 353 353  
 FT DISULFID 153 277  
 FT DISULFID 188 204  
 FT DISULFID 322 338  
 FT DISULFID 349 381  
 FT CARBOHYD 112 112  
 SO SEQUENCE 417 AA; 45011 MW; B2086F661E51D7 CRC64.  
 Query Match 20.8%; Score 565.5; DB 1; Length 417;  
 Best Local Similarity 30.7%; Pred. No. 4,6e-35;  
 Matches 133; Conservative 68; Mismatches 173; Indels 59; Gaps 9;  
 QY 77 CTSKTKKALCTITLTGTFVGAALAGLMLKFMGSKNSGICDSSGICINPSNMCDEV 136  
 Db 12 CCSRPRVYA--ALTFATLTLTAIGA-----ASMAIVA 41  
 QY 137 SHCPGGEDENCVRLYGPNIILQVSSORSMHPVCOODMNNENYGRAACRDGKNNFY 196  
 Db 42 VILRSQDEPLIPYQVSSADARLWVFDKTEGTWLLCSSRNARVALSCBENGFALALH 101  
 QY 197 SGCIYDDSGSTSMKINTAGANDIYK-KLYHSDA-----CSKAVVSLRCAICAV 246  
 Db 102 SELDVATAGA-----NGMSGFCVDEGRPLPHORLEVIYVCDPCRGHFLAICDDCG- 154  
 QY 247 NLNSSRQSRVYGESALPQAMPQVSLHONVHVCSSITTPETIVTAHCVKPLNPHW 306  
 Db 155 -RRKLVDRIYVGRDTSLSGMPQVSLRYDGAHLGSSLSLSDGWLTAACHPRERNRVL 213  
 QY 307 HMTAFAGILROSMFYGAGYQVEKISHPNY-----DSKTKNNDIALMKLOKPLTFNDL 360  
 Db 214 RMVAFGAVAGASP-HGLQGVQVAVYHGGYLPFRPNSEBNSDIALVHSPPLRTEY 272  
 QY 361 VKPVCPLPNPQMLQPEQLCWISGWTGATEKTSVLNAKAVLLIETORCNSRYVDNL 420  
 Db 273 ICPVCLPAAGQALVDKICTVGTGMYQYGOAGVLOEARVPIISNDVCGNADAFYGNQ 332  
 QY 421 TPAMICAGFLQGVDSOCGDSGPLY-----TSKNNITWMLIGDTSMGSCAKAYRPGVYN 476  
 Db 333 KPMFCAGYPEGIDACQDGGSPFVCEDSISRTPRMRLGIVSGTGCALOKRPGVYTK 392  
 QY 477 VAVFTDIYRMR 489  
 Db 393 VSDFRMIFQATK 405  
 RESULT 10  
 ST14\_HUMAN STANDARD; PRT; 855 AA.  
 ID ST14\_HUMAN  
 AC 091516;  
 DT 01-OCT-2000 (rel. 40, Created)

OY 481 TDWYROMRADG 492  
 DB 481 tdwlyrgmrads 492

## RESULT 3

Y92050 Y92050 standard; Protein; 492 AA.

AC Y92050;

DE 01-AUG-2000 (first entry)

DE HRPc6/7 polypeptide from androgen-inducible gene clone.

KM Androgen inducible; testosterone; prostate cancer; cytostatic; TMPRSS2; diagnosis.

OS Homo sapiens.

PN W0200018961-A2.

PD 06-APR-2000.

PF 30-SEP-1999; 99MO-US22535.

PR 30-SEP-1998; 98US-0163759.

PR 30-SEP-1998; 98US-0164159.

PA (MILL-) MILLENNIUM PHARM INC.

PI Macbeth KJ, Shyjan AW;

PI WPI: 2000-293182/25.

DR N-PSDB; A08803.

PT Novel methods for identifying compounds for treating prostate cancer

PT comprising measuring the level of expression or activity of 1 or more

PT of 11 genes or their products

PS Claim 2; Fig 3; 108pp; English.

CC This protein is encoded by a gene which is androgen (e.g. testosterone)

CC and constitutively expressed in androgen-independent prostate cancer

CC cells (e.g. LNCaP cells). Agents which decrease the expression or

CC activity of these clones may slow or arrest the growth of prostate cancer

CC cells or may kill them. HRPc6/7 can be obtained from the sequence of

CC the known gene for TMPRSS2. A compound useful for treating prostate

CC cancer can be identified in a novel method comprising measuring the

CC expression level, or activity, of HRPc6/7, 3, 6/7, 8, 9, 10, 13, 14, 15,

CC 19, or peripheral-type benzodiazepine receptor (PBR) in a cell, in the

CC presence and absence of a test compound. The sequences may also be used

CC in diagnosis of prostate cancer and to determine efficacy of treatment

CC for prostate cancer.

SO Sequence 492 AA;

Query Match 99.7%; Score 2709; DB 21; Length 492;  
 Best Local Similarity 99.8%; Pred. No. 1.1e-202;  
 Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MALNNGSPRAIGPYENHGVQENPYPAQPIVPTVYEVHNAQYPPSPVQYARVLTA 60

DB 1 malnngspalpapyehnygpenyppaqvptvyevehpapypypvayaprvtlga 60

OY 61 SNPVYCQPKSPSGTCTSKTKALCTLTLTGTLVGAALAGLIMFGSKSCNSGTEC 120

DB 61 snpvcqpkspsgtctsktkalcitltltgtlvgalaagllwtfmgskcnsngtec 120

OY 121 DSSGTCINPSNMGDGVSHCPGGEDENRCLVILGPNFLIQVYSQRRKSWHPVCQDDMNEY 180

DB 121 dssgtcinspnmcgvshcpgeedenrcvrlilgpnflilqvysqrrkswhpvcqddmney 180  
 OY 181 GRAACRDMGKYNKFSSQGIYDSDGSTFMRKLTMSAGNVDIYRKLHSDACSSKAVVSLR 240  
 DB 181 graacrdmgyknkfssqglvdsdgsfmrkltmsagnvdiyrklhsdacsksavvslr 240  
 OY 241 CIACGVNLSSROSRIYGESALPGANPKOVSLHVNVAHVGSGSIITPEVITPAHCVEK 300  
 DB 241 ciacgvnlssrsgriygesalpganpkoyslhnvnaahvgsgsiitpevitpahcvek 300  
 OY 301 PLNPMHTAFACILROSEFYGAGYOEKVISHPNVDSTKKNNDIALMKLOKPLFENDL 360  
 DB 301 plnmphtafacilrsefygagyevekvishpnvdstkknnndialmklokpplfendl 360  
 OY 361 VKPYCLPFGMLQPEOLCWIISGCGATEEKGKTEVNAARVLLIETORCNSRYVDNLI 420  
 DB 361 vkpyclpfgmqlqpeqlcwlisgwgateekgkcevlnaarkvllietqrnsryvndl 420  
 OY 421 TPANICAGFLQGNVDSGQSGGPELVTSKNNIMWLIDTISWGSQCAKAPYGVNMF 480  
 DB 421 tpanicagflgnvdsqsggpevltsknnlwlldtswsgcakaypygvnmvf 480  
 OY 481 TDWYROMRADG 492  
 DB 481 tdwlyrgmrads 492

## RESULT 4

ID B36901 B36901 standard; Protein; 492 AA.

AC B36901;

DE 26-FEB-2001 (first entry)

DE Human TMPRSS2 protein.

KM Prostate specific androgen regulated protein; ARSDRL; TMPRSS2;

KM PART-1; neoplastic.

OS Homo sapiens.

PN W0200065067-A2.

PD 02-NOV-2000.

PF 21-APR-2000; 2000MO-US10920.

PR 23-APR-1999; 99US-0130778.

PR 30-AUG-1999; 99US-0151585.

PR 30-DEC-1999; 99US-0174003.

PR 24-JAN-2000; 2000US-0177751.

PA (UNIV ) UNIV WASHINGTON.

PI Nelson PS, Hood L, Lin B;

PI WPI: 2000-679676/66.

DR N-PSDB; C83325.

PT Polynucleotide encoding prostate specific androgen regulated

PT polypeptides and inhibitor of the peptides useful for treating or

PT reducing the progression of prostate neoplastic condition in an

PT individual -

Claim 63; Page 86-88; 121pp; English.

The present invention relates to prostate specific androgen regulated proteins. The invention may be used to determine an expression level of the prostate-specific proteins ARSDRL, TMPRSS2, or PART-1 in a fluid sample or prostate cell sample from an individual. It may also be used for diagnosing and predicting the susceptibility of a

```

RESULT 7
ENTR_PIG STANDARD: PRT: 1034 AA.
AC 01-FEB-1996 (Rel. 33, Created)
BT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRS7 OR ENTRK.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9823;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=DuoDenal mucosa;
RA MEDLINE=94327548; PUBMED=8051081;
RA Matsushima M., Ichinose M., Yanagi N., Kakei N., Tsukada S.,
RA Miki K., Kurokawa K., Tashiro K., Shiohara K., Shimomura K.,
RA Umeiyama H., Inoue H., Takahashi T., Takahashi K.;
RA "Structural characterization of porcine enteropeptidase.";
RA J. Biol. Chem. 269:19976-19982(1994).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN AND TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC TRYPSINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN
CC (HEAVY) CHAIN, AND A MINI CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROPEASE.
CC -1- PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isd-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
EMBL: D30799; BA006459.1; -
DR HSSP: P00763; IDPO.
DR MEROPS: S01.156; -.
DR InterPro: IPR000082; -.
DR InterPro: IPR000859; -.
DR InterPro: IPR000998; -.
DR InterPro: IPR001190; -.
DR InterPro: IPR001254; -.
DR InterPro: IPR001314; -.
DR InterPro: IPR002172; -.
DR Pfam: PF00431; CUB; 2.
DR Pfam: PF00629; MAM; 1.
DR Pfam: PF01390; SEA; 1.
DR Pfam: PF00530; SRCR; 1.
DR Pfam: PF00057; Id1_recept_a; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PRO0722; CHYMOTRYPSIN.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS00740; MAM_1; 1.

```

```

DR PROSITE: PS0060; MAM_2; 1.
DR PROSITE: PS01209; LDLRA_1; 2.
DR PROSITE: PS0068; LDLRA_2; 2.
KV Signal-anchor; Glycoprotein; Myristate; Hydrolase;
KV Serine protease; Zymogen; Transmembrane; Repeat.
FT CHAIN 52 117
FT CHAIN 118 799
FT CHAIN 800 1034
FT TRANSMEM 19 47
FT DOMAIN 197 238
FT DOMAIN 240 349
FT DOMAIN 357 519
FT DOMAIN 539 649
FT DOMAIN 656 694
FT DOMAIN 693 786
FT ACT_SITE 840 840
FT ACT_SITE 891 891
FT ACT_SITE 986 986
FT LIPID 2
FT DISULFID 199 212
FT DISULFID 206 225
FT DISULFID 219 236
FT DISULFID 658 670
FT DISULFID 665 683
FT DISULFID 677 692
FT DISULFID 787 911
FT DISULFID 825 941
FT DISULFID 925 992
FT DISULFID 956 971
FT DISULFID 982 1010
FT CARBOHYD 116 116
FT CARBOHYD 147 147
FT CARBOHYD 170 170
FT CARBOHYD 194 194
FT CARBOHYD 283 283
FT CARBOHYD 333 343
FT CARBOHYD 350 350
FT CARBOHYD 403 403
FT CARBOHYD 455 455
FT CARBOHYD 485 485
FT CARBOHYD 518 518
FT CARBOHYD 549 549
FT CARBOHYD 645 645
FT CARBOHYD 697 697
FT CARBOHYD 701 701
FT CARBOHYD 721 721
FT CARBOHYD 740 740
FT CARBOHYD 761 761
FT CARBOHYD 804 804
FT CARBOHYD 863 863
FT CARBOHYD 902 902
FT CARBOHYD 964 964
SQ SEQUENCE 1034 AA; 114776 MM; 038864CF64CC368 CRC64;

Query Match 23.3%; Score 634; DB 1; Length 1034;
Best Local Similarity 33.3%; Pred. No. 1e-39;
Matches 147; Conservative 72; Mismatches 179; Indels 44; Gaps 13;

OY 65 VCTQPKSPSGIVCTSKTKKALCIT---TLG---TFLVGAALAGLLMKFGSKCSNSG 117
DB 610 VYTGPEVEDVESTNRMTVLEITNDALTKGGFKANFTGYHLG-----IPKCKEDN 662
OY 118 IECDSSGTCINSMNCDGSHCGGEDENRCRLTG--PNFLILOYSSQSKSMHPVCOOD 175
DB 663 FQCE-NGECVLLVNLCDGSHCKDSDEHCVFLNGRANNSGLVQFRIQISIMHTACEN 721
OY 176 WNEENTGRACRDMGYKNN-----FYSSQGIYDSDGTSFMKLNTSAGNWDYKXLYHSDA 230
DB 722 WTTQNSDDVCQLGLGTGNTSSMPFS-----SCGGFVRLNTAPNLSL---LTASEO 771
OY 231 CSKKAVSLRC--IACGVNLNSSROS-RIVGGSALPGAMPVQVSLHVNQVHGGGSLIT 287

```

```

XX 18-JUL-2000 (first entry)
XX Human prostate-associated protease (HUPAP).
DE Human prostate-associated protease; HUPAP; kallikrein; serine protease;
XX gastrointestinal disorder; cancer; prostate disorder.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Modified-site 4 /note="N-glycosylated"
XX Misc-difference 235 /label="unknown
XX /note="Encoded by CNM"
XX
XX US6043033-A.
XX
XX 28-MAR-2000.
XX
XX 27-FEB-1997; 97US-0807151.
XX
XX 27-FEB-1997; 97US-0807151.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Bandman O, Lal P;
XX
XX WPI; 2000-282523/24.
XX
XX N-PSDB; A12975.
XX
XX Polynucleotide encoding human prostate-associated protease useful for
XX diagnosing and treating cancers; prostate disorders and
XX gastrointestinal disorders
XX
XX Claim 1; Fig 1A-B; 27pp; English.
XX
XX This sequence represents human prostate-associated protease (HUPAP). cDNA
XX encoding HUPAP was initially identified in a spinal cord cDNA library,
XX the cDNA encoding this sequence representing a consensus of overlapping
XX and/or extended nucleic acid sequences from spinal cord, prostate tumour
XX and colon cDNA libraries. HUPAP is a serine protease with structural and
XX functional homology with bovine enterokinase, human pancreatic kallikrein
XX and African rat renal kallikrein, sharing 38% homology with bovine
XX enterokinase. In addition, HUPAP is related to prostate-specific antigen
XX (PSA), a kallikrein which is a highly sensitive marker for prostate
XX cancer. HUPAP and nucleic acids encoding it are useful for the diagnosis,
XX prevention and treatment of gastrointestinal disorders such as ulcerative
XX colitis, pancreaticitis, cancers, and prostatic disorders. HUPAP agonists
XX may be used to treat gastrointestinal disorders, and HUPAP antagonists
XX and inhibitors may be used to suppress excessive cell proliferation,
XX which is of use in cancer therapy. HUPAP nucleic acids are also useful
XX for generating hybridisation probes which may be used for mapping
XX naturally occurring genomic sequences.
XX
XX Sequence 283 AA:

```

```

Query Match 56.7%; Score 1540; DB 21; Length 283;
Best Local Similarity 99.6%; Pred. No. 4.2e-112;
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 210 MKNTSAGNDIYKRLVHSAACSKAVYSLRCIAGVNLNNSRSRIVGSGSALPGAMPW 269
DB 1 MKNTSAGNDIYKRLVHSAACSKAVYSLRCIAGVNLNNSRSRIVGSGSALPGAMPW 60
OY 270 QVSLHVNVAHVGCGSITPEMIVTAACVKEPLNNPMHMTAFACILKOSFMFYAGAYVE 329
DB 61 qvslhvnvavcgsiltpeivtaahcveklpnnpmhmtafacilkgfmgaygve 120
OY 330 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKVCPLNPGMLOPEQLCHTISWGATEE 389
DB 1 KVISHPNYDSKTKNDIALMKLQKPLTFNDLVKVCPLNPGMLOPEQLCHTISWGATEE 389

```

```

DB 121 kvishpnysdsktkndialmkqlkpltfndlvkvcplnpgmmlqpeqlcvlsgwatee 180
OY 390 KGTSEVLNAKAVLLIETORNSRYVDNLITPAMICAGFLOGNDVSCGDSGCLPTYSK 449
DB 181 kgtsevlnaakvllietqcnaryyddnlitpamlcagflogndvscgdsqgxlvtak 240
OY 450 NNTWMLIGDTSKSGCAKAVRPGVYGVNMYFTDWIYROMADG 492
DB 241 nntwmligdtswsgcakayrpgvygvnmvftdwlyrqmzdg 283

```

```

RESULT 7
ID B32246 standard; Protein; 454 AA.
XX B32246;
XX
XX 11-JAN-2001 (first entry)
XX
XX Tumour associated differentially-expressed gene 12 protein sequence.
XX
XX Transmembrane serine protease; TADG-12; chromosome 17; vaccination;
XX tumour associated differentially-expressed gene 12; cytosolic; human;
XX malignant hyperplasia; cancer; ovary; breast; lung; colon; prostate.
XX
XX Homo sapiens.
XX
XX MO200052044-A1.
XX
XX 08-SEP-2000.
XX
XX 02-MAR-2000; 2000MO-US05612.
XX
XX 03-MAR-1999; 99US-0261416.
XX
XX (UYAR-) UNIV ARKANSAS.
XX
XX O'Brien TJ, Underwood LJ;
XX
XX WPI; 2000-533263/48.
XX
XX N-PSDB; A93842.
XX
XX DNA fragment encoding tumor associated differentially-expressed gene 12
XX protein used for diagnosing and treating malignant hyperplasia and
XX cancers including ovarian cancer
XX
XX Claim 3; Figure 4; 118pp; English.
XX
XX This invention relates to a novel transmembrane serine protease called
XX tumour associated differentially-expressed gene 12 (TADG-12). TADG-12 is
XX located on chromosome 17. Sequences A93842-A93845 and B32246-B32249
XX represent human TADG-12 cDNA and their corresponding protein sequences. A
XX splice variant of TADG-12 (TADG-12V) leads to a truncated protein
XX product. TADG-12 is overexpressed in ovarian carcinomas. TADG-12 exhibits
XX cytosolic activity, and can be used in vaccines and in gene therapy.
XX TADG-12 nucleotide and protein sequences are used in the diagnosis of
XX malignant hyperplasia and cancers of the ovary, breast, lung, colon,
XX prostate and other cancers where TADG-12 is overexpressed. TADG-12
XX proteins or fragments can be used to vaccinate an individual with cancer,
XX suspected of having a cancer or at risk of getting cancer. Sequences
XX A93846-A93853 represent PCR primers used for amplifying the TADG-12 cDNA
XX sequence, and in the quantitative analysis of TADG-12 mRNA. B32250
XX represents a peptide fragment of TADG-12, used to create anti-TADG-12
XX antibodies. Sequences B32251-B32369 represent TADG-12 peptides which
XX target HLA, and may be used in a vaccine or for immune stimulation.
XX
XX Sequence 454 AA:

```

```

Query Match 32.7%; Score 888; DB 21; Length 454;
Best Local Similarity 46.3%; Pred. No. 3.6e-61;
Matches 190; Conservative 52; Mismatches 148; Indels 20; Gaps 10;

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: U09860; AAC50138.1; -

DR EMBL: Y19124; CAB65555.1; JOINED.

DR EMBL: Y19125; CAB65555.1; JOINED.

DR EMBL: Y19126; CAB65555.1; JOINED.

DR EMBL: Y19127; CAB65555.1; JOINED.

DR EMBL: Y19128; CAB65555.1; JOINED.

DR EMBL: Y19129; CAB65555.1; JOINED.

DR EMBL: Y19130; CAB65555.1; JOINED.

DR EMBL: Y19131; CAB65555.1; JOINED.

DR EMBL: Y19132; CAB65555.1; JOINED.

DR EMBL: Y19133; CAB65555.1; JOINED.

DR EMBL: Y19134; CAB65555.1; JOINED.

DR EMBL: Y19135; CAB65555.1; JOINED.

DR EMBL: Y19136; CAB65555.1; JOINED.

DR EMBL: Y19137; CAB65555.1; JOINED.

DR EMBL: Y19138; CAB65555.1; JOINED.

DR EMBL: Y19139; CAB65555.1; JOINED.

DR EMBL: Y19140; CAB65555.1; JOINED.

DR EMBL: Y19141; CAB65555.1; JOINED.

DR EMBL: Y19142; CAB65555.1; JOINED.

DR EMBL: Y19143; CAB65555.1; JOINED.

DR EMBL: AL163218; CAB90392.1; -

DR HSSP: P00763; IDPO.

DR MEROPS: S01.156; -

DR MIM: 226200; -

DR InterPro: IPR000082; -

DR InterPro: IPR000859; -

DR InterPro: IPR000988; -

DR InterPro: IPR001190; -

DR InterPro: IPR001254; -

DR InterPro: IPR001314; -

DR InterPro: IPR002172; -

DR Pfam: PF00431; CUB; 2.

DR Pfam: PF00629; MAM; 1.

DR Pfam: PF01390; SEA; 1.

DR Pfam: PF00530; SRCR; 1.

DR Pfam: PF00057; Ldl\_recept\_a; 2.

DR Pfam: PF00089; trypsin; 1.

DR PRINTS: PRO0722; CHYMOTRYPSIN.

DR PROSITE: PS00134; TRYPsin\_HIS; 1.

DR PROSITE: PS00135; TRYPsin\_SER; 1.

DR PROSITE: PS01180; CUB; 2.

DR PROSITE: PS00740; MAM\_2; 1.

DR PROSITE: PS01209; LdlRA\_1; 2.

DR PROSITE: PS01209; LdlRA\_2; 2.

DR Signal-anchor: Glycoprotein; Myristate; Hydroxylase; Serine protease; Zymogen; Transmembrane; Repeat.

DR CHAIN 1 784 NON-CATALYTIC CHAIN (HEAVY CHAIN).

FT TRANSMEM 785 1019 CATALYTIC CHAIN (LIGHT CHAIN).

FT DOMAIN 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).

FT DOMAIN 182 223 LDL-RECEPTOR CLASS A.1.

FT DOMAIN 225 334 CUB.

FT DOMAIN 342 504 MAM.

FT DOMAIN 524 634 CUB.

FT DOMAIN 641 679 LDL-RECEPTOR CLASS A.2.

FT DOMAIN 678 771 SRCR.

FT ACT\_SITE 825 825 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 826 826 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT\_SITE 971 971 MYRISTATE (POTENTIAL).

FT LIPID 2 2 BY SIMILARITY.

FT DISULFID 184 197 BY SIMILARITY.

FT DISULFID 191 210 BY SIMILARITY.

FT DISULFID 204 221 BY SIMILARITY.

FT DISULFID 643 655 BY SIMILARITY.

FT DISULFID 650 668 BY SIMILARITY.

FT DISULFID 662 677 BY SIMILARITY.

FT DISULFID 772 896 INTERCHAIN (BY SIMILARITY).

FT DISULFID 810 826 BY SIMILARITY.

FT DISULFID 910 977 BY SIMILARITY.

FT DISULFID 941 956 BY SIMILARITY.

FT DISULFID 967 995 BY SIMILARITY.

FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 440 440 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 134 134 O -> E (IN REF. 3).

FT CONFLICT 732 732 S -> P (IN REF. 3).

FT CONFLICT 754 771 SOOGLDSTLRQCNHKS -> RRNAKKEIDALSPITLIA (IN REF. 3).

SQ SEQUENCE 1019 AA; 112923 MM; B6AA245FD4A563 CRC64;

## Query Match

Best Local Similarity 23.8%; Score 647; DB 1; Length 1019;  
Matches 138; Conservative 67; Mismatches 154; Indels 26; Gaps 10;

OY 113 CSNSGIECDSSGFCINPSMWCDSHCPGDEENRCVRLYG-----PNFIIYSSQ 164  
DB 643 CKADHQ-C-KNECVPVLYNLCDSGLHCEGSDPADCVRFNGTNNGLVRRIO----- 696

OY 165 RKSMPVQDDNMENYGRAACRDMGYKNFYSSOGIVDGSSTFVKLTYSAGNDIYRK 224  
DB 697 -SIMHTACAEENMTTQISNDVCCQLGLGSG-NSSKPIFSTDGC-PFVKLTMAP---DGLI 750

OY 225 LYSDDACSKAVYSLRC--IACGVNLSNR-QSRITGGSALPGAMPVYSLHYGVNHC 281  
DB 751 LTPSQGLDSTLRQCNHKSCKKTLAODITPKYIVGSSNAEGAMPVYVGLYGGRLC 810

OY 282 GGSIIIPENIVTAACVCEPLNPMHTAFAGILROSFPF--YGAGYQYKYSHPNYS 339  
DB 811 GASLVSDMLVSAHCVYGRNLEPSKMTAILGLHMSNLTSPQTVRLDELIVNHYNR 870

OY 340 KTKNDIALMKLQKPLTFENDLYKPVCLPNEGMLLOPEQLCWISGWCATEKGTSEVLA 399  
DB 871 RRKNDIAMHLEFKYNYDYIDPILCPENGVFPGRNCISLAGMCTVVYVQGTANILDE 930

OY 400 AKVLETPQRNCRYYVNDLIPAMICAGFLQGVNDSOGDGGSPLYTSKNINIMLIGT 459  
DB 931 ADVPLLSNENCQOQMPREYN-ITENMICAGIEEGIDSCGDSGGLPMECENRMFLAGT 989

OY 460 -SMGSCAKAYRPGVYGNVAVFTWDI 484  
DB 990 SFGYKCALPNNRGVAVRSFTEMI 1014

## RESULT 6

ENTR\_MOUSE STANDARD: PRT: 1069 AA.

AC P97435;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).  
GN PRSS7 OR ENTK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.



```

FT DOMAIN 151 243 SRCR
FT ACT_SITE 236 492 CATALYTIC
FT ACT_SITE 236 296 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 345 345 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 441 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 113 126 BY SIMILARITY.
FT DISULFID 120 139 BY SIMILARITY.
FT DISULFID 133 148 BY SIMILARITY.
FT DISULFID 244 365 BY SIMILARITY.
FT DISULFID 281 297 BY SIMILARITY.
FT DISULFID 410 426 BY SIMILARITY.
FT DISULFID 437 465 BY SIMILARITY.
FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 492 AA: 53847 MM: 7EAF6DA18609DDA CRC64;

Query Match 99.2%; Score 2696; DB 1; Length 492;
Best Local Similarity 98.8%; Pred.No. 1.3e-193;
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 MALNCGSPPAIGYENHGYQENPYPAQFTVYEVHAPQYSPYQVAPRYLQA 60
DB 1 MALNCGSPPAIGYENHGYQENPYPAQFTVYEVHAPQYSPYQVAPRYLQA 60
OY 61 SNPVCTQPKSPSGVCTSKTKALCITLTGTFVLGALALMLKMGSKCSNGIEC 120
DB 61 SNPVCTQPKSPSGVCTSKTKALCITLTGTFVLGALALMLKMGSKCSNGIEC 120
OY 121 DSSGCTINSNMGDGYSHCGGEGDENRCVRLGPNFILOYSSORSMHPVCCDDMNEY 180
DB 121 DSSGCTINSNMGDGYSHCGGEGDENRCVRLGPNFILOYSSORSMHPVCCDDMNEY 180
OY 181 GRACADMGKKNFYSSQGIYDSDGSGTFMKLTSGAGNDIYKRLYHSDACSSKAVSLR 240
DB 181 GRACADMGKKNFYSSQGIYDSDGSGTFMKLTSGAGNDIYKRLYHSDACSSKAVSLR 240
OY 241 CIACGVNLSSRSRIVGESALPGAMPQVSLHVONVHVGSGSIITPERIVTAACVER 300
DB 241 CIACGVNLSSRSRIVGESALPGAMPQVSLHVONVHVGSGSIITPERIVTAACVER 300
OY 301 PLNPNHMAFAFAGILQSPFYGAGQVEYVISHPNYDSTKKNDAIAMLQPLTFNDL 360
DB 301 PLNPNHMAFAFAGILQSPFYGAGQVEYVISHPNYDSTKKNDAIAMLQPLTFNDL 360
OY 361 VKRYVCLPNEGMLQPOLCISMGATEEKGKSEVILNAKVLIIETQRCNSRYVDNLI 420
DB 361 VKRYVCLPNEGMLQPOLCISMGATEEKGKSEVILNAKVLIIETQRCNSRYVDNLI 420
OY 421 TPAMICAGFLQGNVDSQGDGSGPLVTSKNNIWMLIGDTSWGSQCAKAPRYGNMVF 480
DB 421 TPAMICAGFLQGNVDSQGDGSGPLVTSNNNIWMLIGDTSWGSQCAKAPRYGNMVF 480
OY 481 TDWYIROMRADG 492
DB 481 TDWYIROMRADG 492

RESULT 2
TMS2_MOUSE STANDARD: PRT; 490 AA.
ID TMS2_MOUSE
AC 09J108; 09JUK4; 090Y82;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRANSMEMBRANE PROTEIN, SERINE 2 (EC 3.4.21.-) (EPITHELIALIN) (PLASMIC
GN TRANSMEMBRANE PROTEIN X).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_TakId=10090;
[1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C.
RA Vaarala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT "Expression of transmembrane serine protease Tmprss2 in mouse and
human tissues."
RL J. Pathol. 191:0-0(2000).
RP SEQUENCE FROM N.A.
RA Han J., Kim S.;
RT "Putative transmembrane protease X."
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C.
RA Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
RT "A novel mosaic serine protease, epitheliasin."
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF199362; AAF97867.1; -
CC EMBL: AF243500; AAF64186.1; -
CC EMBL: AF113596; AAF21308.1; -
CC HSSP: P00761; IAKS.
CC MGD: MG1:1354381; Tmprs62.
CC InterPro: IPR001254; -
CC InterPro: IPR001314; -
CC InterPro: IPR002172; -
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC PROSITE: PS01209; LDLRA_1; 1.
CC PROSITE: PS0068; LDLRA_2; 1.
CC Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
CC KW DOMAIN 1 83
CC TRANSMEM 84 104
CC
CC FT DOMAIN 105 490
CC FT DOMAIN 111 149
CC FT DOMAIN 150 242
CC FT DOMAIN 254 490
CC FT ACT_SITE 294 294
CC FT ACT_SITE 343 343
CC FT ACT_SITE 439 439
CC FT DISULFID 112 125
CC FT DISULFID 119 138
CC FT DISULFID 132 147
CC FT DISULFID 243 363
CC FT DISULFID 279 295
CC FT DISULFID 408 424
CC FT DISULFID 435 463
CC FT CARBOHYD 111 111
CC FT CARBOHYD 212 212
CC FT CARBOHYD 474 474
CC FT CARBOHYD 474 474
CC FT CONFLICT 122 122
CC FT CONFLICT 178 178
CC FT CONFLICT 320 320
CC FT CONFLICT 474 474
CC SEQUENCE 490 AA: 53479 MM: 07DB03EAD8A1A9 CRC64;

Query Match 77.9%; Score 2117; DB 1; Length 490;

```







FT	Modified-site	148	/note=	"protein kinase C phosphorylation site"
FT	Modified-site	176	/note=	"N-glycosylated"
FT	Modified-site	197	/note=	"protein kinase C phosphorylation site"
FT	Modified-site	200	/note=	"protein kinase C phosphorylation site"
FT	Modified-site	231	/note=	"protein kinase C phosphorylation site"
FT	Active-site	243	/note=	"casein kinase II phosphorylation site"
FT	Modified-site	249	/note=	"characteristic of serine protease"
FT	Modified-site	260	/note=	"CAMP- and CGMP-dependent protein kinase phosphorylation site"
FT	Modified-site	303	/note=	"protein kinase C phosphorylation site"
FT	Modified-site	351	/note=	"protein kinase C phosphorylation site"
FT	Modified-site	360	/note=	"protein kinase C phosphorylation site"
FT	Modified-site	365	/note=	"tyrosine kinase phosphorylation site"
FT	Modified-site	385	/note=	"protein kinase C phosphorylation site"
FT	Active-site		/note=	"characteristic of serine protease"
XX	WO9936550-A2.			
XX	22-JUL-1999.			
XX	12-JAN-1999;	99WO-US000655.		
XX	16-JAN-1998;	98US-0008271.		
XX	(INCY-)	INCYTE PHARM INC.		
XX	Bandman O, Corley NC, Guejler KJ, Hillman JL, Shah P;			
XX	Tang YT, Yue H;			
XX	WPI: 1999-430616/36.			
DR	N-PSDB: X87154.			
XX	Novel human protease molecules useful in the treatment of			
PT	developmental disorders and/or cancers			
XX	Claim 1; Page 74-75; 90pp; English.			
XX	The present sequence represents novel human protease HUPM-6, as			
CC	deduced from the consensus sequence (see X87154) of overlapping			
CC	cDNA clones obtained from various libraries. Northern analysis			
CC	shows expression of HUPM-6 in gastrointestinal, and male and			
CC	female reproductive cDNA libraries. Approximately 65% of these			
CC	libraries are associated with neoplastic disorders, and 22% with			
CC	the immune response. The invention provides 12 new human			
CC	proteases, i.e. HUPM-1 to -12 (see Y66432-43), and the			
CC	polynucleotides encoding them (see X87149-60). Also provided are			
CC	vectors, host cells and methods for producing HUPM polypeptides,			
CC	as well as agonists and antagonists of HUPM. Methods for treating			
CC	or preventing cell proliferative disorders and immune disorders			
CC	using HUPM or HUPM antagonists are claimed.			
XX	Sequence 435 AA;			
QY	Query Match	24.9%	Score 676.5;	DB 20; Length 435;
DB	Best Local Similarity	39.1%	Pred. No. 9.3e-45;	
	Matches 150; Conservative 57; Mismatches 128;		Indels 49; Gaps 133	

OY		182	RACRDMGY--KNNFYSSQ-----GIYDSSGSTSFVKLTNTSGNVDYIKLKYHSDACS	232
Db		135	etaccmgyskprftraealdppddldlveeltensgeqlimnsq-----pcl	162
OY		233	SKAVVSLERICAGVNLSNRCSRIVGESALPGAMPQVSLHWONHVCGSITTPMIV	292
Db		183	sgslivshlnacgesel---kprvyggeeaaevdwpqvslgydkqhvcgsalldprhwI	239
OY		293	TAAACVEPLNPMPHHWAFACILR-QSPMFYGAGYOYEKVY---SHRWDSKTNNDIAl	348
Db		240	leaahcfirhh-dvdrnmwkragsdsklgaf---palavakllillefnpmY--pkdnlaI	292
OY		349	MKLKPLEFNDLVPCVLPIPNGMALOPEQLCMISGMCATEER-GKTSSEVNAAKVLLIEt	407
Db		292	mklqflptfgstgvrcpiclpffideeltpactplwlvgvftkgngskmsdlilqasvgyIdS	351
OY		408	QRCSNRYVYNLITPAMICAGFLQGNVDSOCDGSGGLPYTSKNINWLIGDTSMGSCAK	467
Db		352	trenaddeaygevetekmmcapipegvdtccgdsgdgrplymqsqd-vhvvglysvwgycCgS	410
OY		468	AAPRGVYGVNWVFETDDMYTRQMRAD	491
Db		411	pstpgvytkvsaylinwlynvwkae	434
RESULT		12		
ID		Y99417		
XX		Y99417	standard; Protein; 432 AA.	
AC		Y99417;		
DT		08-AUG-2000	(first entry)	
DE		Human PRO1570 (UNQ776)	amino acid sequence SEQ ID NO:275.	
KX		Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;		
KW		transmembrane; secretion; immunoadhesion; pharmaceutical; screening.		
KV		Homo sapiens.		
OS		Homo sapiens.		
XX		WO200012708-A2.		
PD		09-MAR-2000.		
XX		01-SEP-1999;	99WC-US20111.	
Pf		01-SEP-1998;	98US-0098716.	
XX		01-SEP-1998;	98US-0098749.	
PR		01-SEP-1998;	98US-0098750.	
PR		02-SEP-1998;	98US-0098803.	
PR		02-SEP-1998;	98US-0098821.	
PR		02-SEP-1998;	98US-0098843.	
PR		09-SEP-1998;	98US-0099336.	
PR		09-SEP-1998;	98US-0099356.	
PR		09-SEP-1998;	98US-0099598.	
PR		09-SEP-1998;	98US-0099602.	
PR		09-SEP-1998;	98US-0099642.	
PR		10-SEP-1998;	98US-0099741.	
PR		10-SEP-1998;	98US-0099754.	
PR		10-SEP-1998;	98US-0099763.	
PR		10-SEP-1998;	98US-0099792.	
PR		10-SEP-1998;	98US-0099808.	
PR		10-SEP-1998;	98US-0099812.	
PR		10-SEP-1998;	98US-0099815.	
PR		10-SEP-1998;	98US-0099816.	
PR		13-SEP-1998;	98US-0100385.	
PR		15-SEP-1998;	98US-0100388.	
PR		15-SEP-1998;	98US-0100390.	
PR		16-SEP-1998;	98US-0100584.	
PR		16-SEP-1998;	98US-0100627.	
PR		16-SEP-1998;	98US-0100661.	
PR		16-SEP-1998;	98US-0100662.	

---

```

Query Match Similarity      32.2%; Score 875.5; DB 1; Length 454;
Best Local Similarity      45.9%; Pred. No. 4.2e-58;
Matches 189; Conservative 53; Mismatches 149; Indels 21; Gaps 11.

.Oy  LTLGLFVLCALAAALAKLLFKMGSKCSNSGISEDSSGCTINPNSMCDGVSHPCGEDENRC 148
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      52 IYIGITALIALALAIOLGIHF---DCSGK-YRORSSFKEICELLARDGVSDDCKDGEDEYRC 107
.Oy  149 VRLXGPNFLOLYSSQSRKSMHPVCODDNNENYGRAACRDMGKNNFYSSQGIYVDSGSTS 208
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      108 VRVGGONAVLQYFTA-ASMKTKSCDDMKMGHYANVACAQLGPF-SYVSSDNLRVSLLEGO 164
.Oy  209 FMKLNTSAGNV---DIYKLYHS----DACSSKAYVSLRCLACGVNLNSRDSRYTGGES 261
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      165 FREEFVS-IDHLDPDKVYALHHSVYVREBCAGHVVYTLQCTCGHGRGYS--SRVYGNM 222
.Oy  262 ALPGAMPPOVSLHYONVNVHCGSITTEPEIVYAAACVEKPLNNPMHMTAFACILROSEFM 321
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      223 SULSGMPQWASIQFGYHLCGSSVITPLMTITTAACV-IDLLPMSWTIQVGLV--SLID 279
.Oy  322 YGA-GYQVEKYVISHPNYDSKTKNNDIALMKLOPLTFENDLVKPVCLPNPGMLQPEQLCW 380
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      280 NPAPHELVEKVIYHSKYPKRRLGNDIALMKLAGPLTFENMIDPVCLPNSSENFPPDGKVCW 339
.Oy  381 ISGCATEE-KGKTSEVLNAKAVLLIEORCSRIVYDNLITPAMICAGFLOGNDSCG 439
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      340 TSGWCATDEGADGADSPVNLHAAPLISNKKICHRDVGIGISPSMCAQYLTVGVDSCG 399
.Oy  440 DSGGFLYTSKNKIMVLIQDTSNGSCAKARYRGVGNVVFWDIYRQNRAD 491
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      400 DSGGFLVCOERLMLKLVGATSTFGICAEYKNGVITRVISFLDWIHEOMERO 451

RESULT 4
ENTR_BOVIN STANDARD; PRT: 1035 AA.
AC P96072.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 03-MAY-2000 (Rel. 33, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRESS OR ENTR.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP TISSUE-Duodenum;
RC MEDLINE=94329561; PubMed=8052624;
RX Kitanoto Y., Yuan X., Wu Q., McCourt D.W., Sadler J.E.,
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RT protease composed of a distinctive assortment of domains.";
RT Proc. Natl. Acad. Sci. U.S.A. 91:7586-7592(1994).
[2]
RN SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE=94043122; PubMed=8226855;
RX Lavallie E.R., Rehentulla A., Racle L.A., Diblasio E.A.,
RA Fernandez C., Grant K.L., Light A., McCoy J.M.;
RA "Cloning and functional expression of a cDNA encoding the catalytic
RT subunit of bovine enterokinase.";
RT J. Biol. Chem. 268:23311-23317(1993).
[3]
RN SEQUENCE OF 801-827.
RP TISSUE-Intestine;
RC MEDLINE=92189715; PubMed=1799406;
RX Light A., Janska H.;
RA "The amino-terminal sequence of the catalytic subunit of bovine
RT enterokinase.";
RT RT J. Protein Chem. 10:475-480(1991).
CC -!- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
PROTEOLYTIC PROENZYMES (TRYPsin, CHYMOTRYPSIN AND CARBOXYPEPTIDASE

```

CC	-1-	CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN TRYPSINOGEN.
CC	-1-	SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC	-1-	SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC	-1-	TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC	-1-	PM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC	-1-	SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC	-1-	SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC	-1-	SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC	-1-	SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC	-1-	SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPsin FAMILY.
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch</a> ).
CC		-----
DR	EMBL;	U09859; AAB40026.1; -
DR	EMBL;	L19663; AAA16035.1; -
DR	PIR;	A61436; A61436.
DR	HSSP;	P00763; IDPO.
DR	MEROPS;	S01.156; -
DR	InterPro;	IPRO000882; -
DR	InterPro;	IPRO00859; -
DR	InterPro;	IPRO000998; -
DR	InterPro;	IPRO01190; -
DR	InterPro;	IPRO01254; -
DR	InterPro;	IPRO01314; -
DR	InterPro;	IPRO02172; -
DR	Pfam;	PF00431; CUB; 2.
DR	Pfam;	PF00629; MAM; 1.
DR	Pfam;	PF01390; SEA; 1.
DR	Pfam;	PF00530; SRCR; 1.
DR	Pfam;	PF00057; Idl_recept_a; 2.
DR	Pfam;	PF00089; CLypsin; 1.
DR	PRINTS;	PR00722; CHYMOTRYPSIN.
DR	PROSITE;	PS00134; TRYPSIN_HIS; 1.
DR	PROSITE;	PS00135; TRYPSIN_SER; 1.
DR	PROSITE;	PS01180; CUB; 2.
DR	PROSITE;	PS00740; MAM_1; 1.
DR	PROSITE;	PS50060; MAM_2; 1.
DR	PROSITE;	PS01209; LDLRA_1; 2.
DR	PROSITE;	PS50068; LDLRA_2; 2.
KM	Signal-anchor:	Glycoprotein; Myristate; Hydrolase.
KV	Serine protease;	Zymogen; Transmembrane; Repeat; Alternative splicing.
FT	CHAIN	1..800
FT	CHAIN	801..1035
FT	TRANSMEM	19..47
FT	DOMAIN	197..238
FT	DOMAIN	240..350
FT	DOMAIN	358..520
FT	DOMAIN	540..650
FT	DOMAIN	657..695
FT	DOMAIN	694..787
FT	ACT_SITE	841..842
FT	ACT_SITE	892..891
FT	ACT_SITE	987..987
FT	LDIFD	2..2
FT	DISULFID	199..212
FT	DISULFID	206..225
FT	DISULFID	219..236
FT	DISULFID	659..671
FT	DISULFID	666..684
FT	DISULFID	678..693

OY 413 RYVDNLITPMICAGFCQGVNDSCGDSGCPLYTSKNNIMWLGDTMSGCAKARPG 472  
 Db 354 ddaygvevtekmkmacglpegvdtcgsdgpplmygsdq-whvvglyvawgygcgspstcp 412  
 OY 473 VYGNMVFETDIYROMRAD 491  
 Db 413 VYKVSAYLWVYVWKAe 431  
 RESULT 13  
 B08950  
 ID B08950 standard; Protein; 480 AA.  
 AC B08950;  
 DT 30-AUG-2000 (first entry)  
 DE Human secreted protein sequence encoded by gene 22 SEQ ID NO:107.  
 KW Human; secreted protein; cytostatic; anti-proliferative; vulnary;  
 KW immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;  
 KW hyperproliferative disorder; infectious disease; tissue regeneration;  
 KW screening; food additive; preservative; wound healing;  
 KW hyper-vascular disease.  
 OS Homo sapiens.  
 PN MO200017222-A1.  
 PD 30-MAR-2000.  
 PE 22-SEP-1999; 99MO-US22012.  
 PR 23-SEP-1998; 98US-0101546.  
 PR 02-OCT-1998; 98US-0102895.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, N1 J;  
 PI Komatsoulis G, Endress GA, Soppet DR;  
 DR WPI: 2000-283538/24.  
 PT Human secreted proteins and coding sequences useful in diagnostic and  
 PT therapeutic methods for disorders such as immune system or  
 PT proliferative disorders, related to the proteins  
 PS Disclosure; Page 65; 416pp; English.  
 CC The polynucleotide sequences given in A39052 to A39088 encode the human  
 CC secreted proteins given in B08891 to B08984. The human secreted proteins  
 CC can have activities based on the tissues and cells they are expressed in.  
 CC Examples of the activities are: cytostatic; anti-proliferative;  
 CC immunosuppressive; antibacterial; and vulnary. The secreted proteins  
 CC and their related polynucleotide sequences are useful for diagnostic and  
 CC therapeutic methods useful for diagnosing and treating disorders related  
 CC to the secreted proteins. The proteins, and polynucleotide sequences may  
 CC be useful for treating disorders of the immune system, hyperproliferative  
 CC disorders, infectious disease, regeneration of tissues, for chemotaxis  
 CC and for screening molecules that bind to the proteins. The proteins or  
 CC polynucleotide sequences may be used as food additives or preservatives,  
 CC to increase or decrease storage capabilities, fat content, lipid,  
 CC protein, carbohydrate, vitamins, minerals, co-factors or other  
 CC nutritional components, agonists or antagonists of the proteins may be  
 CC used to prevent scar tissue growth during wound healing, and  
 CC hyper-vascular diseases. A39043 to A39051 and B08890 are sequences used  
 CC in the exemplification of the present invention.  
 SO Sequence 480 AA;

Best Local Similarity 34.8%; Pred. No. 6,3e-44;  
 Matches 157; Conservative 60; Mismatches 189; Indels 45; Gaps 12;  
 OY 57 LTOASNPVCTOPKSPSGVCTSKKALCTITLIGFVLGALAGLWFMKSGKSN 115  
 Db 45 lsgacp-aaalpyvsfsgavcmrigrcaavlglgyl-llegaygswllvlylcpaasqp 102  
 OY 116 -SGIECDSSGTCINPSMKDGVSHCPGGEDEKRCV-----RLGPNFIILOYYSOR 165  
 Db 103 lsgtlqdeeltl-----scseaaaeallpalpkvtsfrlnsedfllleaayrdq 151  
 OY 166 KSMHPCODDKENMYGRACRDMGY-KNMFYSOGIVDDSGSTFMLNLSAGVNDYRK 224  
 Db 152 pvalllcnegvspalglqlcwslghlrlthkgyvnlcd-----ltnsqefaqdspr 204  
 OY 225 L-YHSDA-----CSSKAVVSLRCIACGVNLNNGSRIVGEGSALPGAMPQVSLHVQ 276  
 Db 205 lgfleeawgprnctagqvslrcsegar---plaslrvgsqavpgrvwpwgsvalg 261  
 OY 277 NVHVCGGSIITPEMIVTAHCEVK-PLNPNHMTAFAGILNQSFNFGAGQVVKVISHP 335  
 Db 262 fhcgsavvlprvvtahcmhsfrlarlsawrvhaglvhsavvrphgvalverllphp 321  
 OY 336 NYDSKTKNDIAIMKIQKPLTFNDLYKPVCLPNPGMMLQPEQLCMISGWSGATEEKG-TS 394  
 Db 322 lvsqnhdydvalllrlqalnfstdvgaclpakqhfpxgrcvcwsgvgnthpshtys 381  
 OY 395 EVLNAAKVLLETORCNRYVDNLITPMICAGFCQGVNDSCGDSGCPLYTSKNNIMW 454  
 Db 382 dmlqdtvplfstqlcnsccvysgaltrpmicaylqgradaegdsdgpplvcpdgdtwr 441  
 OY 455 LIGDTSMGSCGCAKAVRPGVGNVMTDIY 485  
 Db 442 lvgvswgrgaephpyvakvaefldwih 472  
 RESULT 14  
 B11698  
 ID B11698 standard; Protein; 445 AA.  
 AC B11698;  
 DT 23-OCT-2000 (first entry)  
 DE Mouse serine protease BSSP2 (mBSSP2), SEQ ID NO:8.  
 KW BSSP2; serine protease; mouse; mBSSP2; human; hBSSP2; brain;  
 KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;  
 KW epilepsy; cancer; inflammation; inflammatory disorder; infertility;  
 KW prostatic hypertrophy.  
 OS Mus sp.  
 PN MO200031272-A1.  
 PD 02-JUN-2000.  
 PE 19-NOV-1999; 99MO-JP06475.  
 PR 20-NOV-1998; 98JP-0347785.  
 PA (FUSO ) FUSO PHARM IND LTD.  
 PI Uemura H, Okui A, Komitani K, Yamaguchi N, Mitsui S;  
 DR WPI: 2000-400082/34.  
 DR N-PSDB; A61662.  
 PT Serine protease BSSP2, useful in detecting homologs, mutants and  
 PT polymorphic variants as markers for diagnosis of e.g. Alzheimer's  
 PT disease, cancer, inflammation and prostate hypertrophy, using blood,  
 PT urine or other tissues



Tue May 8 11:56:55 2001

us-09-323-597b-2.rag

Page 13

[illegible]

Search completed: May 8, 2001, 11:55:26  
Job time: 46 sec

Qy	332	YGA-GYGVKXISHPNDSTKNDNDILMLQKPLTFNDKPKVCNPNNGMMQLQPEQLCM	380
Db	280	npapshvelkvlyvshkykpkrlgndnlmklapllftfneadgpcipnseenfpgdkvkw	339
Qy	381	ISGWGATEEKGKTSIEVLNAAKVLLIETQRCNSRYVDNLITPAMICAGFLQGNVDSQGD	440
Db	340	tsygvatedggsdaasrvinhaavplnsklcnhvdrygllslpmlcaaylltgvdcagcd	399
Qy	441	SGGFLVSKNNIMWLIGDITSMGSCCAARYPRGYGNMVFDMIRYQMRAD	491
Db	400	sggplvccqerrlmklyvatsfglycaevnkpqytlrvtsflldvlhqeamed	450
RESULT	10		
ID	Y72093	Y72093 standard; Protein: 327 AA.	
XX	Y72093		
AC	XX		
XX	Y72093		
DT	28-MAR-2001	(first entry)	
XX			
DE	Human serine protease #4 encoded by clone HCHAK72.		
XX			
KW	Human; serine protease; osteopathic; immunosuppressive; anti-allergic;		
KW	anti-inflammmatory; cytostatic; cardiant; neuroprotective; nootropic;		
KW	neuroleptic; vulnary; ophthalmological; antibacterial; antiviral;		
KW	antifungal; antiparasitic; gene therapy; diagnosis; prevention; glaucoma		
KW	treatment; bone formation disorder; osteoporosis; arthritis; cancer;		
KW	connective tissue disorder; autoimmune disorder; wound healing; asthma;		
KW	systemic lupus erythematosus; male reproductive system disorder;		
KW	testicular cancer; digestion and food absorption disorder; arrhythmia;		
KW	Crohn's disease; neurodegenerative disease; Alzheimer's disease; allergy		
KW	behavioral disorder; Tourette's syndrome; acute myelogenous leukaemia;		
KW	cardiovascular disorder; ocular disorder; drug screening.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200068247-A2.		
XX			
PD	16-NOV-2000.		
XX			
PF	05-MAY-2000; 2000MO-US12207.		
XX			
PR	07-MAY-1999; 99US-0133239.		
PR	20-MAY-1999; 99US-0135163.		
PR	03-AUG-1999; 99US-0147003.		
PR	09-SEP-1999; 99US-0152935.		
PR	01-NOV-1999; 99US-0162979.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Ruben SM, Shi Y, Young PE, Ni J:		
XX			
DR	WPI: 2000-679799/66.		
XX			
DR	N-PSDB: D02323.		
XX			
PT	New nucleic acid molecules encoding human serine protease polypeptides,		
PT	useful for diagnosis, prevention and/or treatment of disorders e.g.		
PT	osteoporosis, lupus erythematosus and Alzheimer's -		
XX			
PS	Claim 12; Page 280-281; 289pp; English.		
XX			
CC	The present sequence is human serine protease #4 from clone		
CC	HCHAK72 (ATCC Deposit No. PTA27).		
CC	It is used in methods for the diagnosis, prevention and treatment of		
CC	various disorders related to serine protease such as bone formation		
CC	disorders (osteoporosis), connective tissue disorders (arthritis),		
CC	autoimmune disorders (systemic lupus erythematosus), wound healing, male		
CC	reproductive system disorders (testicular cancer), digestion and food		
CC	absorption disorders (Crohn's disease), neurodegenerative diseases		
CC	(Alzheimer's disease), behavioral disorders (Tourette's syndrome),		
CC	behavioral disorders (Tourette's syndrome),		

CC	proliferative and cancerous conditions (acute myelogenous leukemia),
CC	allergic reactions (asthma), cardiovascular disorders (arrhythmia),
CC	ocular disorders (glaucoma), and infectious diseases caused by bacteria,
CC	viruses, fungi or parasites. It is also useful for screening therapeutic
CC	compounds. Serine proteases are used as immunological probes or
CC	polymorphic markers for the identification of chromosomes, cells and
CC	tissues in biological samples, identification of male contraceptive
CC	agents, delivery of compositions to targeted cells expressing a
CC	receptor for serine protease, hybridisation probes and molecular weight
CC	markers. Serine protease nucleic acids are also useful in gene therapy.
CC	Note: The present sequence shown in page 280-281 of sequence listing has
CC	been assigned SEQ ID NO: 14. But the sequence, human serine protease
CC	epidermal growth factor (EGF)-like domain (Y72114) shown in
CC	page 12 is also referred as SEQ ID NO: 14.
CC	
XX	Sequence 327 AA:
XX	
Query Match	27.3%; Score 742.5; DB 21; Length 327;
Best Local Similarity	46.7%; Pred. No. 4.9e-50;
Matches 134; Conservative	45; Mismatches 116; Indels 15; Gaps 8
QY	171 VCDNNENYGRACRDMGYKNNFSSOGIVDDSGSTFMKLTNSAGNV--DIYKLYH 227
Db	1 mcsddckkyhyanvacagqigfp-syvsadnllyvslegqfreesidhllpdkvvtalh 59
QY	228 S---DACSSKAVYSLRCIACGVNINSSRSHVGGESALPGAMPQVSLHYONVHYCG 283
Db	60 syvvtgscasghvvtlqctacghrzgys--srlyvgmmsllsqpwqasalfqgylhcg 117
QY	284 SITTPETVTAHCEKPLNPNMHTARFGILROSEFMETGA-GIQVEKVISHPYDSKTK 342
Db	118 svtlpwtlitaahcv-ydilylpkswtclqgylv--sllndpaphllveklvyhskypkrl 174
QY	343 NNIDIALMKLOKPLTFNDLVKPYCLPNPGMLQPEOLCISGSGATEE-KGKTSEVLNAAK 401
Db	175 gndlatmklaaprltfhemqgvclpnsenfpdykvkwtsgtgcagdaapvlnhaa 234
QY	402 VLLIETORCNSRYVYDNLITTPAMICAGFLOGNVDSQCGDSGGLPVTSKNNIMWLIGDTSW 461
Db	235 vplsmklnchrdvvgllspamlcaaylftgsvdsqcgsgplvcqerrllwkvgetsf 294
QY	462 GSGCAKAYRPQVYGVNWFDTNTRYQMRAD 491
Db	295 glgcaevnkpvytrvtcsfldwlhegmard 324







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2001, 11:54:40 ; Search time 20.73 Seconds

(without alignments)  
1631.051 Million cell updates/sec

Title: US-09-323-597B-2

Perfect score: 2717

Sequence: 1 MALNCGSPPAIGPYENHGY.....VYGNVNFPTDITRQMRADG 492

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries.

Database :  
1: PIR1:  
2: PIR2:  
3: PIR3:  
4: PIR4:

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	660	24.3	1035	1 A43090	enteropeptidase (E
2	647	23.8	1019	1 A56318	enteropeptidase (E
3	634	23.3	1034	1 A53663	enteropeptidase (E
4	577.5	21.3	1113	2 JE0315	low-density lipopr
5	569.5	21.0	638	1 KQHUP	plasma kallikrein
6	565.5	20.8	417	1 S00845	hepsin (EC 3.4.21.
7	556	20.5	638	1 KORPPL	plasma kallikrein
8	538	19.8	638	1 KOMSPL	plasma kallikrein
9	533	19.6	625	1 KFHU1	coagulation factor
10	529.5	19.5	416	1 S33777	hepsin (EC 3.4.21.
11	518.5	19.1	821	1 PLM6	plasma (EC 3.4.21
12	501	18.4	421	1 S11674	acrosin (EC 3.4.21
13	501	18.4	1524	2 T30337	polypeptidase - Afri
14	490.5	18.1	415	1 A34170	acrosin (EC 3.4.21
15	489.5	18.0	436	2 JX0172	acrosin (EC 3.4.21
16	486.5	17.9	761	2 JCS759	brain-specific ser
17	483	17.8	418	2 A37344	acrosin (EC 3.4.21
18	477	17.6	790	1 PLPG	plasma (EC 3.4.21
19	475.5	17.5	431	2 S47538	acrosin (EC 3.4.21
20	473.5	17.4	421	2 S29599	acrosin (EC 3.4.21
21	472.5	17.4	810	2 B30848	plasma (EC 3.4.21
22	470.5	17.3	437	2 S18407	acrosin (EC 3.4.21
23	470.5	17.3	810	2 I46260	plasma (EC 3.4.21
24	461	17.0	343	1 A47014	proctasin (EC 3.4.
25	456.5	16.8	460	2 B61545	plasma (EC 3.4.21
26	455	16.7	812	1 PLBO	plasma (EC 3.4.21
27	451	16.6	2616	2 A57096	nudel protein prec
28	450.5	16.6	267	2 S40006	trypsin (EC 3.4.21
29	450	16.6	270	2 S56160	mast cell tryptase

30	449	16.5	275	2 S40005	trypsin (EC 3.4.21
31	449	16.5	655	1 A46688	hepatocyte growth
32	448	16.5	277	2 S35340	trypsin (EC 3.4.21
33	447	16.5	276	2 A38654	mast cell proteinase
34	446.5	16.4	420	2 A55283	acrosin (EC 3.4.21
35	445.5	16.4	810	1 PLHU	plasma (EC 3.4.21
36	444.5	16.4	455	2 A61545	plasma (EC 3.4.21
37	443.5	16.3	266	2 S54146	trypsin (EC 3.4.21
38	442.5	16.3	4548	1 S00657	apoptin (a) (EC
39	441.5	16.2	274	2 JC4171	trypsin (EC 3.4.2
40	441	16.2	786	1 A47547	serine proteinase
41	439	16.2	273	2 A47246	trypsin (EC 3.4.2
42	438	16.1	274	2 S35339	trypsin (EC 3.4.21
43	437	16.1	275	2 S40007	trypsin (EC 3.4.21
44	436	16.0	242	2 S49489	trypsin (EC 3.4.21
45	436	16.0	456	1 KXBO	protein C (activat

## ALIGNMENTS

RESULT 1  
A43090  
enteropeptidase (EC 3.4.21.9) precursor - bovine  
N:Alternate names: enterokinase  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A43090; A48874; A61436  
R:Kitamoto, Y.; Yuan, X.; Mu, Q.; McCourt, D.W.; Sadler, J.E.  
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994  
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease co  
A:Reference number: A43090; MUID:94329561  
A:Accession: A43090  
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1035 <KIR>  
A:Cross-references: GB:U09859; NID:9746410; PIDN:AA0026.1; PID:9746411  
A:Experimental source: small intestine  
R:Lavallie, E.R.; Rehentulla, A.; Reicle, L.A.; DiBlasio, E.A.; Ferez, C.; Grant, K.L  
J. Biol. Chem. 268, 23311-23317, 1993  
A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit o  
A:Reference number: A48874; MUID:94043122  
A:Accession: A48874  
A:Molecule type: mRNA  
A:Residues: 801-1035 <LAV>  
A:Cross-references: GB:L19663; NID:9416131; PIDN:AA16035.1; PID:9416132  
A:Note: parts of this sequence, including the amino end of the mature protein, were c  
R:Light, A.; Janska, H.  
J. Protein Chem. 10, 475-480, 1991  
A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.  
A:Reference number: A61436; MUID:92189715  
A:Accession: A61436  
A:Molecule type: protein  
A:Residues: 801-807, 'Y', 809-827 <LIC>  
C:Comment: The mechanism of association with the membrane of the intestinal brush bor  
embrane attachment using a signal-anchor sequence.  
C:Comment: Conversion from membrane-bound to soluble forms may involve further proces  
C:Complex: mature enteropeptidase is variously reported to contain two (heavy and lig  
liffe linked  
C:Function:  
A:Description: cleaves propeptide from trypsinogen to produce active trypsin  
A:Pathway: intestinal digestive hydrolase cascade  
C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding  
C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane prot  
F:32-18/Domain: transmembrane #status predicted <TM>  
F:32-117/Product: enteropeptidase mini chain #status predicted <KCH>  
F:118-800/Product: enteropeptidase heavy chain #status predicted <KCH>  
F:199-236/Domain: LDL receptor ligand-binding repeat homology <KCH>  
F:358-520/Domain: MAM homology <MAM>  
F:542-647/Domain: Clr/Cls repeat homology <CLR>  
F:659-693/Domain: LDL receptor ligand-binding repeat homology <CLR>  
F:694-799/Domain: scavenger receptor cysteine-rich domain homology <LDR2>  
F:801-1035/Product: enteropeptidase light chain #status predicted <LCH>

F:801-1030/Domain: trypsin homology <TRY>  
 F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Bindin  
 F:788-912,826-842,926-933,957-972,983-1011/Disulfide bonds: #status predicted  
 F:841,892,987/Active site: His, Asp, Ser #status predicted

Query Match 24.3%; Score 660; DB 1; Length 1035;  
 Best Local Similarity 33.9%; Pred. No. 5e-40;  
 Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;

65 VCTQPKSPSGTGVCTSKTKKALCITLT-----GTELVGAALAGLIMKEMSKCSNSG 117  
 DB 611 VYRGPVNVDFSTNMTWLFITDNMLAKQGFANNTTGVGLG-----IPEPCEDN 663  
 QY 118 IEEDSSSTCINPSNCDGVSHCPGEGDENRCVRLYG-----PNTLYVSSQKSMH 169  
 DB 664 FQC-KQDECIPLVNLCDGFPHCKDGSDEAHCVRLFNTTDSGLVPRIO-----SIWH 716  
 QY 170 PVODDMENYGRACRDMQYKNNFYSSQGIYDSDGSTSPMKLNTSAGNDIYKILHSD 229  
 DB 717 VACAEHTTQISDDVLCOLLGLGTG--NSVPTSTGGGPVNLTAENGSLT---LTPSQ 771  
 QY 230 ACSSKAVYSLRC--IACGVNLNNSROS-RIVGESALPGAMPQVSLHYQNVHVGSGII 286  
 DB 772 QCLDEDLILLOCHYKSCGKLTQEVSPKIVGSDSREGAMPVVALYFDQVCGASLV 831  
 QY 287 TTPWYITAAHCVEKPLNPNHMTAFAGILROSEMFYAGYQVE-----KVISHPNYSKT 341  
 DB 832 SRMLVSAACHVGRNNEPSPKRAVLELHNASNL--TSPQIEIRLLIDQVINYHVKRR 888  
 QY 342 KNNIDIALMKLOKPLTENDLVKPCLPNPGMLOPEOLCWTISGMCATEBEKTSFVLNAK 401  
 DB 889 KNNIDIAMHLEKVNITDYIOTICLPENOVFPFGKICSIAGKALYQGSTADYLDQAD 948  
 QY 402 VLLIETQRNSRYVDNLITPAMICAGFLQGVNDSCGSDGSPLYTSKNNIMWILIGDTW 461  
 DB 949 VPLLSNERCQOQMEYN-ITENMYCAGYEAGVDSGCGSGPLMCCENRMWLAGVTSE 1007  
 QY 462 GSGCAKAYRPGVYGNVAFETDWT 484  
 DB 1008 GYOCALPNRPGVYARVRFTEWI 1030

## RESULT 2

enteropeptidase (EC 3.4.21.9) precursor - human

N:Alternate names: enterokinase  
 C:Species: Homo sapiens (man)  
 C>Date: 19-MAY-1995 #sequence\_revision 09-Aug-1996 #text\_change 18-Jun-1999

R:Kilamoto, Y.; Velle, R.A.; Donis-Keller, H.; Sadler, J.E.  
 Biochemistry 34, 4562-4568, 1995

A:Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolyt  
 A:Reference number: A56318; MUID:95234679

A:Accession: A56318

A:Molecule type: mRNA

A:Residues: 1-1019 <KIT>

A:Cross-references: GB:U09860; NID:q746412; PIDN:AAC50138.1; PID:q746413

R:Kilamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994

A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo  
 A:Reference number: A43090; MUID:94329561

A:Accession: B43090

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 749-1019 <K12>

A:Cross-references: GB:U09860

C:Comment: The mechanism of association with the membrane of the intestinal brush border  
 (noted below) or with amino-terminal myristoylation of the heavy chain.

C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and lig  
 ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms inv  
 ducts.

C:Function: cleaves activation peptide from trypsinogen to produce active trypsin

A:Pathway: Intestinal digestive hydrolase cascade

C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding

C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymoge

F:1-784/Product: enteropeptidase heavy chain #status predicted <HCH>

F:184-221/Domain: transmembrane #status predicted <TM>

F:342-504/Domain: NAM homology <NAM>

F:526-631/Domain: Clr/Cls repeat homology <CLR>

F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <

F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>

F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Bindin

F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted

F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 23.8%; Score 647; DB 1; Length 1019;

Best Local Similarity 35.8%; Pred. No. 4.3e-39;

Matches 138; Conservative 67; Mismatches 154; Indels 26; Gaps 10;

QY 113 CANSIGECDSSTCINPSNCDGVSHCPGEGDENRCVRLYG-----PNTLYVSSQ 164  
 DB 643 CKADHQC-KNGECVPLVNLCDHLCEGSDPADCVRFNCTNNGLVRRIO----- 696  
 QY 165 RSMHPVPCDDNMENTGRAACRDMQYKNNFYSSQGIYDSDGSTSPMKLNTSAGNDIYK 224  
 DB 697 -STWHRAEAMNTTQISDDVLCOLLGLGTG--NSKPLFSTDCG-PFKWLTAP---DGLI 750  
 QY 225 LYHSDACSSKAVYSLRC--IACGVNLNNSR-OSRIVGESALPGAMPQVSLHYQNVHVC 281  
 DB 751 LTPSQOCLDLSLIRLOCNHKSCKKLAADIPPKIYGGSNAGKAMPVVALYFDQVCGRLC 810  
 QY 282 GGSITTPWYITAAHCVEKPLNPNHMTAFAGILROSEMFYAGYQVE-----KVISHPNYSKT 339  
 DB 811 GASLVSDMLVSAACHVGRNNEPSPKRAVLELHNASNL--TSPQIEIRLLIDQVINYHVKRR 870  
 QY 340 KTKNNIDIALMKLOKPLTENDLVKPCLPNPGMLOPEOLCWTISGMCATEBEKTSFVLNA 399  
 DB 871 RKNNDIAMHLEKVNITDYIOTICLPENOVFPFGKICSIAGKALYQGSTADYLDQAD 930  
 QY 400 AKVLLIETQRNSRYVDNLITPAMICAGFLQGVNDSCGSDGSPLYTSKNNIMWILIGDT 459  
 DB 931 ADVPLLSNERCQOQMEYN-ITENMICAGYEBEGIDSCGSDGSGPLMCCENRMWLAGVT 989  
 QY 460 SWGSCAKAYRPGVYGNVAFETDWT 484  
 DB 990 SFGYKCALPNRPGVYARVRFTEWI 1014

## RESULT 3

enteropeptidase (EC 3.4.21.9) precursor - pig

N:Alternate names: enterokinase

C:Species: Sus scrofa domestica (domestic pig)

C>Date: 07-Oct-1994 #sequence\_revision 09-Aug-1996 #text\_change 18-Jun-1999

R:Matsumura, M.; Ichinose, M.; Yahagi, N.; Kakai, N.; Tsukada, S.; Miki, K.; Kurokawa

J. Biol. Chem. 269, 19976-19982, 1994

A:Title: Structural characterization of porcine enteropeptidase.

A:Reference number: A53663; MUID:94327548

A:Accession: A53663

A:Molecule type: mRNA

A:Residues: 1-1034 <MAT>

A:Cross-references: GB:D30799; NID:9505122; PIDN:BA06459.1; PID:9505123

C:Note: parts of this sequence, including the amino ends of three chains isolated fro

C:Comment: The mechanism of association with the membrane of the intestinal brush bor

(noted below) or with amino-terminal myristoylation of the heavy chain.



Inogen and may also play a role in the renin-angiotensin system by converting prorenin I  
 C:Genetics:  
 A:Gene: GDB:KLR3  
 A:Cross-references: GDB:127575; OMIM:229000  
 A:Map position: 4q35-4q35  
 C:Superfamily: coagulation factor XI; trypsin homology  
 C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla  
 F:1-19/Domain: signal sequence #status predicted <Sto>  
 F:20-638/Product: plasma kallikrein #status predicted <MAT>  
 F:20-390/Domain: plasma kallikrein heavy chain #status predicted <HCH>  
 F:20-109/Domain: apple repeat <AP1>  
 F:110-139/Domain: apple repeat <AP2>  
 F:200-289/Domain: apple repeat <AP3>  
 F:291-380/Domain: apple repeat <AP4>  
 F:391-638/Domain: plasma kallikrein light chain #status predicted <LCH>  
 F:21-104/Domain: trypsin homology <TRY>  
 F:21-104/47-77/51-57/111-194/137-166/141-147/201-284/227-256/231-237/292-375/322-328/383  
 F:127-308/336/453/494/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:318-347/340-345/Disulfide bonds: #status predicted  
 F:390-391/Cleavage site: Arg-1le (coagulation factor XIa) #status predicted  
 F:434/483/578/Active site: His, Asp, Ser #status predicted

Query Match 21.0%; Score 569.5; DB 1; Length 638;  
 Best Local Similarity 32.8%; Pred. No. 1e-33;  
 Matches 155; Conservative 64; Mismatches 160; Indels 93; Gaps 23;

55 RVLQASNVNVTQTPSPGTVCTSKKALCTLTGLFVLGALAAAGLKKFMKMS--- 111  
 218 RVLVT-PDAPVCC-----TICTYHPN--CLEFTFY-----NVMKIESQRNV 255  
 112 ---KCSNSGIECDG-----SG---TCINP-----SNKCDVSHCPGEGEDENRCVRL 151  
 256 CLTKTSESTPSSSTPQENTISGLTLCKRLRPERCHKITVGVDF--GGELN----- 308  
 152 YGPNFTLQVYSSORSKWHPVCCODDMNENTGRAACRDMGKNNFYSSCGIYDGSSTS--- 208  
 309 --VTFKGV-----NVCQE-----TCTRM-IRCOFFTSLPREDCEKCKC 347  
 209 FKLKNSAANDVIYKKLYSDACSSKAVVSLRCIAGVNV--LNSSQSRVSGESALPGA 266  
 348 FRLKSDGSPTR--AYTQGS--YSLRCNTGDSVCTTKSTRVVGTSNCKE 401  
 267 WPMOVSIAHQ---NVAVCGSITTPEMIVTAACVE--KPLNPMHTATAGILROSEMEY 322  
 402 WPMOVSIAHQVLAQRLLCGSLIGHQVLAHAFGLPGLDQVNR--IYSGILNLSDIRK 459  
 323 GAGY-OVERVISHPNYDSTKKNDAIMLKLQKPLTFNDLVKPYCLPDPGMMQLEQLCWT 381  
 460 DTFPSQIKELIIHONKVKSEGNHDIKLAQAPLNTYEFQKPICLPSKDPSTIYTNQWV 519  
 382 SGMGATDEEKGKSEVANAQVLLIETORNSRNVNLTTPATICGFLQGNVDSOGGS 441  
 520 TWMGFEKKEGELONLQKVNIPLVLTNECOKR--QYKTKTQRRVACVKEGKGADCKGDS 578  
 442 GGPLYTSKNNIMLLIGTSGSCAKARPGVGNMVFETDIYROMR-ADG 492  
 579 GGPLYCKNKGMRVLVITSGECCARREPQVYTKAEYMDWILEKTOSSDG 630

RESULT 6  
 S00845  
 hepsin (EC 3.4.21.-) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 18-Jun-1999  
 C:Accession: S00845  
 R:Levy, S.P.; Loebl, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.  
 Biochemistry 27, 1067-1074, 1988  
 A:Title: A novel trypsin-like serine protease (hepsin) with a putative transmembrane dom  
 A:Reference number: S00845; MUID:88209431  
 A:Accession: S00845  
 A:Molecule type: mRNA  
 A:Residues: 1-417 <LEV>

A:Cross-references: EMBL:X077732; MUID:932063; PIDN:CAA30558.1; PID:932064  
 C:Genetics:  
 A:Gene: GDB:HPN; IMPRSS1; hepsin  
 A:Cross-references: GDB:135685; OMIM:142440  
 A:Map position: 19q11-19q13.2  
 C:Superfamily: hepsin; trypsin homology  
 C:Keywords: hydrolase; liver; serine protease; transmembrane protein  
 F:23-45/Domain: transmembrane #status predicted <TM>  
 F:163-400/Domain: trypsin homology <TRY>  
 F:188-204/291-359/322-338/349-381/Disulfide bonds: #status predicted  
 F:203/257/353/Active site: His, Asp, Ser #status predicted

Query Match 20.8%; Score 565.5; DB 1; Length 417;  
 Best Local Similarity 30.7%; Pred. No. 1.2e-33;  
 Matches 133; Conservative 68; Mismatches 173; Indels 59; Gaps 9;

77 CSKTKKALCTLTGLFVLGALAAAGLKKFMKMSGSGSSTGCTIPSWCCGV 136  
 12 CSRPVVA--ALTAGTLLTLTAIGA-----ASMAIVA 41  
 137 SHCPGEGEDENRCVRLGPNFTLQVYSSORSKWHPVCCODDMNENTGRAACRDMGKNNFY 196  
 42 VLLRSDPELYVQVVSADARLNVFDKTEGTTRLLCSSNARVAGLSCDEKGFRLALTH 101  
 197 SGIYDSSGTSFEMKLNSTAGNVDYK-KLYHSDA-----CSSKAVVSLRCIAGV 246  
 102 SELDVTATGA-----NGISGFVCYDEGLRPHQRLLEYISVDCPRGFLAICODCG- 154  
 247 NUNSSQSRVSGESALPGAHPVOVSLAQNVHVCSSITTEPMIVTAACVEKRLNPMW 306  
 155 -RRKLVPYDITVIGRSTSLRRMPVOVSLRYDGAHLGSSLSLSDWVLAHACPEENRVLS 213  
 307 HMTAFAGILROSEMEYAGVQVEYVISHPNY-----DSKKNDAIMLKLQKPLTFNDL 360  
 214 RRVVTRGAVQAQSP-HGLDQVQAVYHGGYLPFRDPNSENENDIALVHSSPLREY 272  
 361 VVPCLPNDPMALQEPOLCWSMGATEERKGTSEYLAQVLLIETORNSRNVYDNL 420  
 273 IQPVCLPAAGALVQKICTVGVGMVQYGGQAGVLDGARVILSNVDCNADPEYQDI 332  
 421 TPAMTACAGLQGNVDSOGDSGPLY---TSKNNIMLLIGTSGSCAKARPGVGNV 476  
 333 KPEKMGAGYPEGIDACODSGPFVCEDSISRTPRMRLCGIVSWGTGALAKQGVYTK 392  
 477 VVETDIYROMR 489  
 393 VSDFERMIFQAIK 405

RESULT 7  
 K0RTP  
 plasma kallikrein (EC 3.4.21.34) precursor - rat  
 M:Alternate names: Fletcher factor; kininogenin; serum kallikrein  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 18-Jun-1999  
 R:Beauchamp, G.; Rosinski-Chupin, I.; Matted, M.G.; Mdkay, M.; Chretien, M.; Seldah.  
 Biochemistry 30, 1628-1635, 1991  
 A:Title: Gene structure and chromosomal localization of plasma kallikrein.  
 A:Reference number: A39180; MUID:91129236  
 A:Accession: A39180  
 A:Molecule type: DNA  
 A:Residues: 1-638 <BEA>  
 A:Cross-references: GB:J05315  
 A:Note: the authors translated the codon GAG for residue 81 as Gln  
 R:Seldah, N.G.; Lademheim, R.; Mdkay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; La  
 DNA 8, 563-574, 1989  
 A:Title: The cDNA structure of rat plasma kallikrein.  
 A:Reference number: A33320; MUID:90091743  
 A:Accession: A33320  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA



PR	31-MAR-1998;	98US-0080107.
PR	31-MAR-1998;	98US-0080165.
PR	31-MAR-1998;	98US-0080199.
PR	01-APR-1998;	98US-0080327.
PR	01-APR-1998;	98US-0080328.
PR	01-APR-1998;	98US-0080333.
PR	01-APR-1998;	98US-0080334.
PR	08-APR-1998;	98US-0081049.
PR	08-APR-1998;	98US-0081070.
PR	08-APR-1998;	98US-0081071.
PR	09-APR-1998;	98US-0081195.
PR	09-APR-1998;	98US-0081203.
PR	09-APR-1998;	98US-0081229.
PR	15-APR-1998;	98US-0081817.
PR	15-APR-1998;	98US-0081838.
PR	15-APR-1998;	98US-0081952.
PR	15-APR-1998;	98US-0081955.
PR	21-APR-1998;	98US-0082566.
PR	21-APR-1998;	98US-0082700.
PR	22-APR-1998;	98US-0082704.
PR	22-APR-1998;	98US-0082804.
PR	23-APR-1998;	98US-0082767.
PR	23-APR-1998;	98US-0082796.
PR	27-APR-1998;	98US-0083336.
PR	28-APR-1998;	98US-0083332.
PR	29-APR-1998;	98US-0083392.
PR	29-APR-1998;	98US-0083495.
PR	29-APR-1998;	98US-0083496.
PR	29-APR-1998;	98US-0083499.
PR	29-APR-1998;	98US-0083500.
PR	29-APR-1998;	98US-0083544.
PR	29-APR-1998;	98US-0083554.
PR	29-APR-1998;	98US-0083558.
PR	29-APR-1998;	98US-0083559.
PR	30-APR-1998;	98US-0083742.
PR	05-MAY-1998;	98US-0084366.
PR	06-MAY-1998;	98US-0084414.
PR	06-MAY-1998;	98US-0084441.
PR	07-MAY-1998;	98US-0084598.
PR	07-MAY-1998;	98US-0084600.
PR	07-MAY-1998;	98US-0084627.
PR	07-MAY-1998;	98US-0084637.
PR	07-MAY-1998;	98US-0084639.
PR	07-MAY-1998;	98US-0084640.
PR	07-MAY-1998;	98US-0084643.
PR	13-MAY-1998;	98US-0085323.
PR	13-MAY-1998;	98US-0085338.
PR	13-MAY-1998;	98US-0085339.
PR	15-MAY-1998;	98US-0085373.
PR	15-MAY-1998;	98US-0085579.
PR	15-MAY-1998;	98US-0085580.
PR	15-MAY-1998;	98US-0085582.
PR	15-MAY-1998;	98US-0085689.
PR	15-MAY-1998;	98US-0085697.
PR	15-MAY-1998;	98US-0085700.
PR	15-MAY-1998;	98US-0085704.
PR	18-MAY-1998;	98US-0086023.
PR	22-MAY-1998;	98US-0086392.
PR	22-MAY-1998;	98US-0086414.
PR	22-MAY-1998;	98US-0086430.
PR	22-MAY-1998;	98US-0086486.
PR	28-MAY-1998;	98US-0087099.
PR	28-MAY-1998;	98US-0087106.
PR	30-JUL-1998;	98US-0087200.
PR	11-SEP-1998;	98US-0094651.
PR		98US-0100038.
XX		
PA	(GETH ) GENENTECH INC.	
XX		
PI	Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J	
XX		
DR	WPI; 1999-551358/46.	



C:Accession: A38514; S48202; S48203  
 R:Degeen, S.J.F.; Bell, S.M.; Schaefer, L.A.; Elliott, R.W.  
 Genomics 8, 49-61, 1990  
 A:Title: Characterization of the CDNA coding for mouse plasminogen and localization of  
 A:Reference number: A38514; MUID:91184812  
 A:Accession: A38514  
 A:Molecule type: mRNA  
 A:Residues: 1-812 <DEG>  
 A:Cross-references: GB:J04766; NID:J200402; PIDN:AAA50168.1; PID:J200403  
 R:Lijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.  
 Eur. J. Biochem. 224, 863-871, 1994  
 A:Title: Characterization of the murine plasma fibrinolytic system.  
 A:Reference number: S48202; MUID:95010076  
 A:Accession: S48202  
 A:Molecule type: Protein  
 A:Residues: 20-25 <LIJ>  
 A:Accession: S48203  
 A:Molecule type: Protein  
 A:Residues: 22-27 <LIJ>  
 C:Comment: Plasminogen is synthesized by the kidney and is present in plasma and many of  
 C:Comment: plasminogen is converted into plasmin by plasminogen activators, both plasmin  
 immediately after dissociation from the clot. In the presence of the inhibitor, the active  
 C:Comment: inhibitor, the activation involves also removal of the activation peptide.  
 C:Comment: Streptolysin 1 (see PIR:KCMS1) acts on plasminogen to produce angiotensin. It  
 etol in treating solid tumors.  
 C:Function:  
 A:Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a va  
 as the walls of the graafian follicle; also activates the urokinase-type plasminogen act  
 A:Pathway: fibrinolysis  
 C:Superfamily: plasmin; kringe homology; plasminogen-related protein precursor homology  
 C:Keywords: angiotensin inhibitor; blood; duplication; fibrinolysis; glycoprotein; hyd  
 F:1-96/Domain: plasminogen-related protein precursor homology <PLPH>  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-812/Product: plasminogen #status predicted <PRO>  
 F:20-96/Domain: activation peptide #status predicted <APT>  
 F:79-466/Product: angiotensin #status predicted <AST>  
 F:79-581,582-812/Product: plasmin #status predicted <AMT>  
 F:97-581/Domain: chain A #status predicted <ACH>  
 F:103-181/Domain: kringe homology <KR1>  
 F:185-262/Domain: kringe homology <KR2>  
 F:275-352/Domain: kringe homology <KR3>  
 F:377-454/Domain: kringe homology <KR4>  
 F:481-560/Domain: kringe homology <KR5>  
 F:582-812/Domain: chain B #status predicted <BCH>  
 F:582-805/Domain: trypsin homology <TRY>  
 F:49-73,53-61,103-124,152-176,185-262,188-316,206-245,234-257,275-352,296-335,32  
 bonds: #status predicted  
 F:78-79/Cleavage site: Glu-Asn (streptolysin 1) #status predicted  
 F:136,308/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:466-467/Cleavage site: Thr-Val (plasminogen activator) #status predicted  
 F:581-582/Cleavage site: Arg-Val (plasminogen activator) #status experimental  
 F:624,667,762/Active site: His, Asp, Ser #status predicted

Query Match 19.1%; Score 518.5; DB 1; Length 812;  
 Best Local Similarity 32.7%; Pred. No. 6; Be-30;  
 Matches 136; Conservative 54; Mismatches 143; Indels 83; Gaps 16;

QY 106 WKPMGSK-CNSGTECDSSGTCI-----NPSMCGVSHNC-PGGEDEKRCVRLGPN 155  
 DB 446 WEYCENLRKCEGT-----GSVELLPVSOEPSPSSEIDCMXGNGKRGKTAVTAG 499  
 QY 156 FTLQVSSQKSMWHPVC-----QDDMNENYGRAACRDMG-----YKNNFYSSQGIYVDSG 205  
 DB 500 TPOGMAAOEPHRHSITTPQTNPRADLEKNYCRNRPDDVDVNPMTYTN-----547  
 QY 206 STFSFMKINTSAGNDYIKKLYHSDAGSSKAVVSRCTACGVNLSSNQ-----SRIVGESAL 263  
 DB 548 -----PRKLY--DYCDIPLCASASFECCGKPVQEPKCKPGRVYVGGCVAN 589  
 QY 264 PGAMPQVSLHVQ--NVHVGGSITTEPMIVTAAHCEKPLNNMWHMTAFAGILROSEMF 321  
 DB 590 PHSMPOISLIRRTGTHGFCGTLIAPEWVLTAAHCEKS--SRPEFYKVLGAHE-----644

QY 322 YGAGTQVE-----KVHPNYSKTKNNIDALMKLQPLFNFNDLYKVCPLNPMQLOPE 376  
 DB 645 YIRLDVQVEISVAKLIEPN-----NRDIALKLSHPATIDVIVACPLSPMYVADR 698  
 QY 377 QLCWISGMGATE---EKGKSEVLANAKVLLIEFROGNSRVYNDLTPMIGCGFLOGN 433  
 DB 699 TICITTMGEETGFGGRLEK---AOLPYENKVCNREYVNNRKSSTELCAGOLAG 754  
 QY 434 VDSQCGSGGLVYSKNNIMLLIDTSMGSCAKAPGYGVNMYFTDMYRQMR 489  
 DB 755 VDSQCGSGGLVYCFEKKDXYILGCVTSMGLGCAKPNKPGYVRSRVDMIERMR 810

RESULT 12  
 S11674  
 acrosin (EC 3.4.21.10) precursor - human  
 C:Species: Homo sapiens (hmn)  
 C:Date: 10-Sep-1999 #sequence, revision 10-Sep-1999 #text, change 16-Jun-2000  
 C:Accession: S11674; S23499; S12063; A61022; S03330  
 R:Keime, S.; Adam, I.M.; Engel, W.  
 Eur. J. Biochem. 190, 195-200, 1990  
 A:Title: Nucleotide sequence and exon-intron organization of the human proacrosin gen  
 A:Reference number: S11674; MUID:90306003  
 A:Accession: S11674  
 A:Molecule type: DNA  
 A:Residues: 1-421 <KEI>  
 A:Cross-references: EMBL:X54017; NID:J93582; PIDN:CAA37964.1; PID:G1216165  
 A:Note: the authors translated the codon AGG for residue 64 as Thr and CAG for residu  
 R:Vazquez-Levin, M.H.; Reventos, J.; Gordon, J.W.  
 Eur. J. Biochem. 207, 23-26, 1992  
 A:Title: Molecular cloning, sequencing and restriction mapping of the genomic sequenc  
 A:Reference number: S23499; MUID:92331659  
 A:Accession: S23499  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-421 <VAZ>  
 A:Cross-references: EMBL:M77378  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992  
 R:Keime, S.  
 submitted to the EMBL Data Library, December 1989  
 A:Reference number: S12063  
 A:Accession: S12063  
 A:Molecule type: DNA  
 A:Residues: 1-225, 'R', 227-421 <KEI2>  
 A:Cross-references: EMBL:X54017  
 R:Adam, I.M.; Klemm, U.; Maier, W.M.; Engel, W.  
 Hum. Genet. 84, 125-128, 1990  
 A:Title: Molecular cloning of human proacrosin cDNA.  
 A:Reference number: A61022; MUID:90128988  
 A:Accession: A61022  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-63, 'T', 65-225, 'V', 227-267, 'R', 269-421 <ADH>  
 R:Baba, T.; Watanabe, K.; Kashiwabara, S.I.; Arai, Y.  
 FEBS Lett. 244, 296-300, 1989  
 A:Title: Primary structure of human proacrosin deduced from its cDNA sequence.  
 A:Reference number: S03330; MUID:89153568  
 A:Accession: S03330  
 A:Molecule type: mRNA  
 A:Residues: 1-63, 'T', 65-110, 'V', 121-165, 'L', 167-267, 'R', 269-344, 'R', 346-421 <BAB>  
 A:Cross-references: EMBL:Y00970; NID:J28325; PIDN:CAA68784.1; PID:J28326  
 C:Genetics:  
 A:Gene: GDB:ACR  
 A:Cross-references: GDB:119645; OMIM:102480  
 A:Map position: 22q13-22qter  
 A:Introns: 26/2; 94/2; 189/2; 237/3  
 C:Superfamily: acrosin; trypsin homology  
 C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-421/Product: acrosin #status predicted <MAT>  
 F:20-44/Product: acrosin light chain #status predicted <LCH>  
 F:43-421/Product: acrosin heavy chain #status predicted <HCH>



RESULT 14  
US-08-451-932-1

Sequence 1, Application US/08451932

Patent No. 5733876

GENERAL INFORMATION:

APPLICANT: Folkman, Judah

APPLICANT: O'Reilly, Michael

TITLE OF INVENTION: Method of Treating an Angiogenic

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Jones &amp; Askew

STREET: 191 Peachtree Street, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303-1769

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50

COMPUTER: Macintosh

OPERATING SYSTEM: 7.0

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/451,932

FILING DATE: 05/26/95

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/248,629

FILING DATE: 04/26/94

ATTORNEY/AGENT INFORMATION:

NAME: Larry W. Stults, Ph.D.

REGISTRATION NUMBER: 34,025

REFERENCE/DOCKET NUMBER: 05213-0123

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-818-3700

TELEFAX: 404-818-3799

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 812

TYPE: amino acid

TOPOLOGY: linear

US-08-451-932-1

Query Match

Best Local Similarity 19.1%; Score 518.5; DB 1; Length 812;

Matches 136; Conservative 54; Mismatches 143; Indels 83; Gaps 16;

DB 106 WKPMGSK-CSNSGIEDSSGTCI-----NPSNMGDGVSHC--PGGEDENRCVRLXGPN 155

DB 446 WEYCNLKRCEGT-----GSVELPTVSGPSPDSETDCKYNGKDYRGKTAATAAG 499

DB 156 FIIQVYSSQKSHNPPC-----QDDMNENYGRAACRDG-----YKNNFYSSOGIVDDSG 205

DB 500 TPQGNMAAOEPRHSHSTPTPTNPRADLEKYNCRPDGVDVGPWCYTTN-----547

DB 206 STSFMKLNTSAGNVDIKKLYHSDACSSKAVSLRCIACGVNLSSRQ--SRIVGSAL 263

DB 548 -----PRKLY--DYCDIPLCASAASSFECKPQVEPRKCGRRVYVGCYVAN 589

DB 264 PGAMPQVSLHVG--NVHVGGSITTPETVITTAHCEYKPLNPMWMTAFAGILRSFME 321

DB 590 PHSMPQOISLRTFTGQHFGLIAPENVLTAAHCLKS-SREFYKVIILGAHEE----644

DB 322 YGAGYQVE-----KVISHPNYDSKTKNDIALMKLQPLTFNDLVKVCILPFGMALQPE 376

DB 645 YIRGLDQVLEISVAKLILERN-----NRDIALKLKSRPATITDKVAPACLPSPNMYADR 698

DB 377 QLCWISGWGATE---EKGKTSEVLNAAKVLIIETORCNRSRYVYDNLITPAMICAGFLQGN 433

DB 699 TIIQYITMGSTGCTGFGGRKE-----AQLPVIENKVCNRRVEYLNNRKSKSELCAQGLAG 754

DB 434 VDSGQDSGGPLVTSKNNIWLIGDTSWGSCKAKAYRPGVYGVNMFETDWTYRQMR 489

DB 755 VDSGQDSGGPLVCFEKKDYIILOGVTSWGSCKARPNKPGVYVRSRFFVMIEREMR 810

RESULT 15  
US-08-452-260-1

Sequence 1, Application US/08452260

Patent No. 5776704

GENERAL INFORMATION:

APPLICANT: Folkman, Judah

APPLICANT: O'Reilly, Michael

TITLE OF INVENTION: Method of Diagnosing an Angiogenic

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Jones &amp; Askew

STREET: 191 Peachtree Street, 37th Floor

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303-1769

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50

COMPUTER: Macintosh

OPERATING SYSTEM: 7.0

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/452,260

FILING DATE: 05/26/95

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/248,629

FILING DATE: 04/26/94

ATTORNEY/AGENT INFORMATION:

NAME: Larry W. Stults, Ph.D.

REGISTRATION NUMBER: 34,025

REFERENCE/DOCKET NUMBER: 05213-0124

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-818-3700

TELEFAX: 404-818-3799

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 812

TYPE: amino acid

TOPOLOGY: linear

US-08-452-260-1

Query Match

Best Local Similarity 19.1%; Score 518.5; DB 1; Length 812;

Matches 136; Conservative 54; Mismatches 143; Indels 83; Gaps 16;

DB 106 WKPMGSK-CSNSGIEDSSGTCI-----NPSNMGDGVSHC--PGGEDENRCVRLXGPN 155

DB 446 WEYCNLKRCEGT-----GSVELPTVSGPSPDSETDCKYNGKDYRGKTAATAAG 499

DB 156 FIIQVYSSQKSHNPPC-----QDDMNENYGRAACRDG-----YKNNFYSSOGIVDDSG 205

DB 500 TPQGNMAAOEPRHSHSTPTPTNPRADLEKYNCRPDGVDVGPWCYTTN-----547

DB 206 STSFMKLNTSAGNVDIKKLYHSDACSSKAVSLRCIACGVNLSSRQ--SRIVGSAL 263

DB 548 -----PRKLY--DYCDIPLCASAASSFECKPQVEPRKCGRRVYVGCYVAN 589

DB 264 PGAMPQVSLHVG--NVHVGGSITTPETVITTAHCEYKPLNPMWMTAFAGILRSFME 321

DB 590 PHSMPQOISLRTFTGQHFGLIAPENVLTAAHCLKS-SREFYKVIILGAHEE----644

DB 322 YGAGYQVE-----KVISHPNYDSKTKNDIALMKLQPLTFNDLVKVCILPFGMALQPE 376

DB 645 YIRGLDQVLEISVAKLILERN-----NRDIALKLKSRPATITDKVAPACLPSPNMYADR 698

DB 377 QLCWISGWGATE---EKGKTSEVLNAAKVLIIETORCNRSRYVYDNLITPAMICAGFLQGN 433



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2001, 11:55:30 ; Search time 11.45 Seconds  
(without alignments)  
1471.939 Million cell updates/sec

Title: US-09-323-597B-2

Perfect score: 2717  
Sequence: 1 MALNSGSPAIQPIYENHGY.....VYGNVMTFTIYROMRADG 492

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0-5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2696	99.2	492	1 TMS2_HUMAN	O15393 homo sapien
2	2117	77.9	490	1 TMS2_MOUSE	O91488 mus musculu
3	875.5	32.2	454	1 TMS3_HUMAN	P57727 homo sapien
4	660	24.3	1035	1 ENTK_BOVIN	P98072 bos taurus
5	647	23.8	1019	1 ENTK_HUMAN	P98073 homo sapien
6	638	23.5	1069	1 ENTK_MOUSE	P97445 mus musculu
7	634	23.3	1034	1 ENTK_PIG	P98074 sus scrofa
8	569.5	21.0	638	1 KAL_HUMAN	P03952 homo sapien
9	563.5	20.8	417	1 HEP5_HUMAN	P03951 homo sapien
10	558.5	20.6	855	1 ST14_HUMAN	O91546 homo sapien
11	556	20.5	638	1 KAL_RAT	P14272 rattus norv
12	548.5	20.2	902	1 ST14_MOUSE	P56677 mus musculu
13	538	19.8	638	1 KAL_MOUSE	P26262 mus musculu
14	536.5	19.7	416	1 HEP5_MOUSE	O35453 mus musculu
15	533	19.6	625	1 FAL1_HUMAN	P03951 homo sapien
16	529.5	19.5	416	1 HEP5_RAT	P03951 homo sapien
17	518.5	19.1	812	1 PLMN_MOUSE	P20918 mus musculu
18	501	18.4	421	1 ACRO_HUMAN	P10323 homo sapien
19	490.5	18.1	415	1 ACRO_PIG	P08001 sus scrofa
20	489.5	18.0	436	1 ACRO_MOUSE	P23578 mus musculu
21	486.5	17.9	761	1 NCTR_MOUSE	O08762 mus musculu
22	477	17.6	790	1 PLMN_PIG	P06867 sus scrofa
23	475.5	17.5	431	1 ACRO_RABIT	P48038 oryctolagus
24	472.5	17.4	810	1 PLMN_MACMU	P12545 macaca mula
25	470.5	17.3	437	1 ACRO_RAT	P29293 rattus norv
26	470.5	17.3	810	1 PLMN_BIEBU	O29465 erinaceus e
27	461	17.0	343	1 PSS8_HUMAN	O16651 homo sapien
28	456.5	16.8	343	1 PLMN_SHEEP	P81286 ovis aries
29	455	16.7	812	1 PLMN_BOVIN	P06868 bos taurus
30	452	16.6	875	1 NCTR_HUMAN	P56730 homo sapien
31	451	16.6	267	1 NDL_DROME	P98159 drosophila
32	450.5	16.6	267	1 TRYT_ANOGA	P35041 anopheles g
33	450	16.6	270	1 TRYT_MERON	P50342 meriones un

34	449	16.5	275	1 TRYT_ANOGA	P35038 anopheles g
35	449	16.5	655	1 HGFA_HUMAN	O04756 homo sapien
36	448	16.5	277	1 TRYT_ANOGA	P35036 anopheles g
37	447	16.5	276	1 MCT6_MOUSE	P21845 mus musculu
38	445.5	16.4	810	1 PLMN_HUMAN	P00747 homo sapien
39	444.5	16.4	338	1 PLMN_HORSE	P80010 equus caball
40	443	16.3	274	1 TRYT_ANOGA	P35035 anopheles g
41	442.5	16.3	4548	1 APOA_HUMAN	P08519 homo sapien
42	441.5	16.2	274	1 TRYT_RAT	P50343 rattus norv
43	441	16.2	653	1 HGFA_MOUSE	O91098 mus musculu
44	441	16.2	786	1 STUB_DROME	O05319 drosophila
45	439	16.2	273	1 MCT7_MOUSE	O02844 mus musculu

## ALIGNMENTS

RESULT 1	STANDARD	PRT	492 AA.
TMS2_HUMAN			
ID TMS2_HUMAN			
AC O15393;			
DT 15-JUL-1998 (Rel. 36, Created)			
DT 15-JUL-1998 (Rel. 36, Last sequence update)			
DT 01-OCT-2000 (Rel. 40, Last annotation update)			
DE TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-).			
GN TMPSR52.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE-97468144; PubMed-9325052;			
RA Polonit-Giacobino A., Chen H., Peltsch M.C., Rossier C.,			
RA Antonarakis S.E.;			
RT Cloning of the TMPSR52 gene, which encodes a novel serine protease			
RT with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3.;			
RL Genomics 44:309-320(1997).			
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.			
CC -1- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE AND			
CC WEAKLY IN SEVERAL OTHER TISSUES.			
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			
CC TRYPSIN FAMILY.			
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.			
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.			
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC the European Bioinformatics Institute. There are no restrictions on its			
CC use by non-profit institutions as long as its content is in no way			
CC modified and this statement is not removed. Usage by and for commercial			
CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC -----			
CC EMBL: U75329; AAC51784.1; ..			
CC HSP: P00763; IDPO.			
CC MIM: 602060;			
CC InterPro: IPR001254; ..			
CC InterPro: IPR001314; ..			
CC InterPro: IPR002172; ..			
CC Pfam: PF00057; IdL_recept_a; 1.			
CC Pfam: PF00089; trypsin; 1.			
CC PRINTS: PR00722; CHYMOTRYPSIN.			
CC PROSITE: PS00134; TRYPSIN_HIS; 1.			
CC PROSITE: PS00135; TRYPSIN_SER; 1.			
CC PROSITE: PS01209; LDLRA_1; 1.			
CC PROSITE: PS50068; LDLRA_2; 1.			
CC Hydrolase; serine protease; Transmembrane; Signal-anchor.			
CC HYDROLASE; serine protease; Transmembrane; Signal-anchor.			
CC CYTOPLASMIC (POTENTIAL).			
CC SIGNAL-ANCHOR (POTENTIAL).			
CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)			
CC (POTENTIAL).			
CC EXTRACELLULAR (POTENTIAL).			
CC LDL-RECEPTOR CLASS A.			
FT DOMAIN 106 492			
FT 112 150			

XX The invention provides a new tumour suppressor gene, designated TMPRSS2.  
 CC The TMPRSS2 polynucleotides and polypeptides can be used in methods for  
 CC diagnosing and prognosis predisposition to cancer in humans. The  
 CC polypeptides may also be used in assays to screen for compounds with  
 CC anti-cancer or therapeutic properties. The polypeptides are also useful  
 CC for rational drug design. The TMPRSS2 polynucleotides and polypeptides  
 CC may be used for gene therapy and protein therapy. The present sequence  
 CC represents the TMPRSS2 polypeptide.

XX Sequence 492 AA:

Query Match 100.0%; Score 2717; DB 21; Length 492;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-203;  
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALSSGSPPAIGPYENHGYOPENPYPAOPTVPTVEVHPAOPYSPPOVAPRYVTOA 60  
 DB 1 malssgppaigpyenhyqpenypaqptvptvevhpapypqyaprvltga 60  
 QY 61 SNPVYCTQPKSPSGTCTVCTSKTKKALCTTLTGLTFLVGAALAGLMMKFMGSKCSNSGIEC 120  
 DB 61 snpvyctqpkspsgtctvctsktkkalcitltlgtflvgaaalagllwmfmgskcsngilec 120  
 QY 121 DSGGTCINPSNMCDCVSHCPGEGEDENRCVRLYGNFTLQVYSSORKSHPVCCDDMNENT 180  
 DB 121 dsggtcinpnmcdvshcpggedenrcvrlgpnftlqvysorkshpvcddmneny 180  
 QY 181 GRACRDMGKKNRNSOGIVDSDGSTSPMKLNTSAGNVDTYKRLYHSDACSSKAVVSLR 240  
 DB 181 gracrdmgyknnlyssqglvddsgstsfmklntsagnvdykrllyhsdaasskavvslr 240  
 QY 241 CIACGVNLSSROSRIYGGESALPGAMPQVSLHVQNVHVCSSGITPEWIVTAHCVCK 300  
 DB 241 ciacgvnlssrsriyggosalpgampqvslhvqnvhvcsgsiltpevltahcvek 300  
 QY 301 PLNNPWHMTAFAGILROSFMFYAGIOVEKVIHNPNTSKTKNDIALMKLQKPLTFNDL 360  
 DB 301 plnnpwhmtafagilrsgfmfyagiovekvihsnpntsktkndialmkqlkpltfndl 360  
 QY 361 VKPVCILNPGMWLOPEOLCMTSGMGATEEKGKSTSEVINAADVLLIETQRCNSRYVDNLI 420  
 DB 361 vkpvcilnpgmwlopolcmtsgmgateekgstsevinadvllietqrcnsryvndli 420  
 QY 421 TPAMICAGFLQGNVDSGCGSGPLVTSKNNIMWLIGDTSMGSCCAKAYRPVYGNMVF 480  
 DB 421 tpamicagflqgnvdsqgsgplvtsknnimwlligdtswgscakayrpvygnmvf 480  
 QY 481 TDWYIRQMRADG 492  
 DB 481 tdwlyrqmrady 492

RESULT 2

Y44406 Y44406 standard; Protein; 492 AA.

XX Y44406;  
 AC Human 20P1F12-GTC2 protein.  
 DT 22-MAR-2000 (first entry)  
 XX DE  
 XX Human 20P1F12-GTC2 protein.  
 XX 20P1F12; TMPRSS2; androgen; serine protease; 20P1F12-GTC1; cancer;  
 KM transmembrane protein; colon; prostate; prostate tumour.  
 XX Homo sapiens.  
 OS  
 XX  
 XX NC099629.42-A2.  
 XX  
 PD 09-DEC-1999.  
 XX

PF 01-JUN-1999; 99MO-US12253.  
 XX  
 PR 01-JUN-1998; 98US-0087598.  
 PR 29-JUN-1998; 98US-0091474.  
 PR 14-APR-1999; 99US-0129521.

XX (UROG-) UROGENESIS INC.  
 PA (AFAR/) AFAR D E.  
 PA (HUBE/) HUBERT R S.  
 PA (LEON/) LEONG K.  
 PA (RAIT/) RAITANO A B.  
 PA (SAFE/) SAFERAN D C.

PI Afar DE, Hubert RS, Leong K, Raitano AB, Saffran DC;  
 XX  
 XX WPI; 2000-116363/10.  
 DR N-PSDB; 229636.

PT Novel cell surface antigen useful to treat colon and prostate cancer -  
 PS Claim 1; Fig 1; 58pp; English.

CC The present sequence is the 20P1F12 protein (also known as the TMPRSS2  
 CC protein) which is a prostate-specific, androgen-regulated, cell surface  
 CC serine protease. It is a glycosylated type II transmembrane protein with  
 CC an extracellular C-terminal serine protease domain, a scavenger receptor  
 CC cysteine-rich domain, an LDL receptor class A domain and a predicted  
 CC transmembrane domain. Host cells can be transformed to produce this  
 CC protein, using vector containing 20P1F12/TMPRSS2 gene (also designated  
 CC 20P1F12-GTC1, as deposited with ATCC accession number 207097).  
 CC Anti-20P1F12/TMPRSS2 antibodies may be used as therapeutic agent for  
 CC prostate and colon cancers, to image prostate cancer cells and  
 CC prostate tumours, to identify ligands and cellular constituents that  
 CC bind to a 20P1F12/TMPRSS2 gene product and for use as cancer vaccines.

SO Sequence 492 AA:

Query Match 100.0%; Score 2717; DB 21; Length 492;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-203;  
 Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALSSGSPPAIGPYENHGYOPENPYPAOPTVPTVEVHPAOPYSPPOVAPRYVTOA 60  
 DB 1 malssgppaigpyenhyqpenypaqptvptvevhpapypqyaprvltga 60  
 QY 61 SNPVYCTQPKSPSGTCTVCTSKTKKALCTTLTGLTFLVGAALAGLMMKFMGSKCSNSGIEC 120  
 DB 61 snpvyctqpkspsgtctvctsktkkalcitltlgtflvgaaalagllwmfmgskcsngilec 120  
 QY 121 DSGGTCINPSNMCDCVSHCPGEGEDENRCVRLYGNFTLQVYSSORKSHPVCCDDMNENT 180  
 DB 121 dsggtcinpnmcdvshcpggedenrcvrlgpnftlqvysorkshpvcddmneny 180  
 QY 181 GRACRDMGKKNRNSOGIVDSDGSTSPMKLNTSAGNVDTYKRLYHSDACSSKAVVSLR 240  
 DB 181 gracrdmgyknnlyssqglvddsgstsfmklntsagnvdykrllyhsdaasskavvslr 240  
 QY 241 CIACGVNLSSROSRIYGGESALPGAMPQVSLHVQNVHVCSSGITPEWIVTAHCVCK 300  
 DB 241 ciacgvnlssrsriyggosalpgampqvslhvqnvhvcsgsiltpevltahcvek 300  
 QY 301 PLNNPWHMTAFAGILROSFMFYAGIOVEKVIHNPNTSKTKNDIALMKLQKPLTFNDL 360  
 DB 301 plnnpwhmtafagilrsgfmfyagiovekvihsnpntsktkndialmkqlkpltfndl 360  
 QY 361 VKPVCILNPGMWLOPEOLCMTSGMGATEEKGKSTSEVINAADVLLIETQRCNSRYVDNLI 420  
 DB 361 vkpvcilnpgmwlopolcmtsgmgateekgstsevinadvllietqrcnsryvndli 420  
 QY 421 TPAMICAGFLQGNVDSGCGSGPLVTSKNNIMWLIGDTSMGSCCAKAYRPVYGNMVF 480  
 DB 421 tpamicagflqgnvdsqgsgplvtsknnimwlligdtswgscakayrpvygnmvf 480

	Best Local Similarity	78.4%	Pred. No. 1,7e+150:	
	Matches	389;	Conservative	41; Mismatches 63; Indels 2; Gaps
OY	1	MALNSGSPRAIGPYENHGYQEPENPYPAQTPVPTVEYEHVPAQIYSPVQVAPRVLTQA	60	
Db	1	MALNSGSPPGIGPCYEHNHGYOHEHICPPRPVPANGLNLYPAQIYPSPVQVAPRITQA	60	
OY	61	SNPVCYQPKRPSSTCTVCTSKTKRALCTTLTLCFTLVGAALAAAGLIMKFMKSCKSNSGITEC	120	
Db	61	STVIHHHPKS-SCAPCTSKSKSLCTALATGLVTGAATAAVALMRFWDNSTSEMEC	119	
OY	121	DSSTCTINPSNMKGVSCHCPGDEBENCVRLYGNFETLOYYSSORKSMHRPCODDMNEY	180	
Db	120	GSGGTCTISSLMCGGVACHPCPBEDBNCCVALLQGSEFLLOYYSORKNMYTCDDMSERY	179	
OY	181	GRACAKDMDGYNKFYSOSGIYVDGSGTSEFKMLNTSAGNVDIYKKLYHSDACSSKAVYSLR	240	
Db	180	GRACKDKMGYNKFYSOSGIYDPDGATSEFKMLNVSNGNDVLTKLYHSDCSSRMVYSLR	239	
OY	241	CICACGVNLNSRGSRIRYGGESALBPAMPNOVSLHVQVAHYHCSSGITPEMTIVTAHCEYK	300	
Db	240	CIECGVR-SVKRKRIYTGGLNASGDMPQVSLHVQVHYHCSSGITPEMTIVTAHCEE	298	
OY	301	PLNNPMHTAAGILTRDSPMFYGXGYOVEKYISHPNVDSKTNNNDIAMLKOLPIAFENDL	360	
Db	299	PLSGRWTAAGILTRDSLFYFSRHQVEKYISHPNVDSKTNNNDIAMLKQLPIAFENDL	358	
OY	361	VKPCLPRLPQGMILLDPECWISMGANGAYEKGKTSDEVLAALMAPLEIPSKCNKRYTNLI	418	
Db	359	VKPCLPRLPQGMILLDDECWISMGANGAYEKGKTSDEVLAALMAPLEIPSKCNKRYTNLI	418	
OY	421	TPAMICAGFLOGANDSCOGSDGPELVYTSKNNIMWLIDTISWGSCCAKAIRPGYGNVMF	480	
Db	419	TPAMICAGFLOGSYDVSCOGSDGPELVTLKNGIMWLIDTISWGSCCAKALPRGYGNVMF	478	
OY	481	TDWIYROWRAD	491	
Db	479	TDWIYQOMRAN	489	

  

RESULT	3
TM53_HUMAN	
ID	TM53_HUMAN
AC	P57727
DT	01-OCT-2000 (Rel. 40, Created)
DT	01-OCT-2000 (Rel. 40, Last sequence update)
DT	01-OCT-2000 (Rel. 40, Last annotation update)
DE	TRANSMEMBRANE PROTEASE, SERINE 3 (EC 3.4.21.-) (SERINE PROTEASE TAG-12) (TUMOR ASSOCIATED DIFFERENTIALLY EXPRESSED GENE-12 PROTEIN).
GN	TPMRSS3 OR TADG12 OR ECHO51.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_TaxID=9606;
NB	[1]
RN	SEQUENCE FROM N.A. (ISOFORMS A AND TRUNCATED).
RP	TISSUE=Ovarian carcinoma;
RC	MEDLINE=20521358; PubMed=11068177;
RA	Underwood L.J., Shigemasa K., Tanimoto H., Beard J.B., Schneider E.N., Wang Y., Paruley T.H., O'Brien T.J.;
RT	"Ovarian tumor cells express a novel multi-domain cell surface serine protease.";
RL	Biochim. Biophys. Acta 1502:337-350(2000).
RN	[2]
RP	SEQUENCE FROM N.A. (ISOFORMS A; B/C AND D).
RC	PubMed=1137999;
RA	Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Christ R., Guptoni M., Wand J., Kawasaki K., Asakawa S., Minoshima S., Younis F., Mehdi S.O., Radhakrishna U., Papasavvas M.P., Gehrig C., Roussier C., Korostitshevsky M., Gal A., Shimizu N., Bonne-Tamir B., Antonarakis S.E.;
RT	"Insertion of beta-satellite repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal

FT	RECESSIVE DEAFNESS."
RL	Nat. Genet. 27:59-63(2001).
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC	-1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; A (SHOWN HERE), B/C, D AND
CC	TRUNCATED/MDG-IV; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. THE TRUNCATED
CC	ISOFORM IS FOUND AT INCREASED LEVELS IN SOME CARCINOMAS.
CC	-1- DISEASE: DEFECTS IN TRPSN3 ARE A CAUSE OF TWO FORMS OF AUTOSOMAL
CC	NEUROSENSORY CHILDHOOD-ONSET FORMS OF DEAFNESS, DENB8 AND DENB10.
CC	-1- SIMILARITY: BELONGS TO PERTIDASE FAMILY S1; ALSO KNOWN AS THE
CC	TRPSIN FAMILY.
CC	-1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC	-1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; AF201380; AAC37012.1; -
DR	EMBL; AB038157; BAB20077.1; -
DR	EMBL; AB038158; BAB20078.1; -
DR	EMBL; AB038159; BAB20079.1; -
DR	EMBL; AB038160; BAB20080.1; -
DR	MIM; 605511; -
DR	MIM; 601072; -
DR	MIM; 605316; -
DR	PROSITE; PS01209; LDLRA_1; 1
DR	PROSITE; PS50068; LDLR_2; 1
DR	PROSITE; PS50287; SRCR_2; 1
DR	PROSITE; PS00134; TRPSIN_HIS; 1
DR	PROSITE; PS00135; TRPSIN_SER; 1
DR	Hydrolase; Serine protease; Transmembrane; signal-anchor; Deafness;
KM	Alternative splicing.
FT	DOMAIN 1 28
FT	TRANSMEM 29 49
FT	DOMAIN 50 454
FT	DOMAIN 72 108
FT	DOMAIN 109 205
FT	DOMAIN 217 454
FT	ACT_SITE 257 257
FT	ACT_SITE 304 304
FT	ACT_SITE 401 401
FT	DISULFID 73 85
FT	DISULFID 79 98
FT	DISULFID 92 107
FT	DISULFID 207 324
FT	DISULFID 242 258
FT	DISULFID 397 425
FT	CARBOHYD 221 221
FT	VARSPLIC 1 127
FT	VARSPLIC 318 454
FT	-----
FT	VARSPPLIC 261 293
FT	VARSPPLIC 294 454
FT	CONFLICT 46 54
FT	CONFLICT 90 90
FT	CONFLICT 350 350
FT	CONFLICT 369 395
FT	CONFLICT 427 427
FT	CONFLICT 454 AA; 49404 MW; 57ECG36787DBA9F CRC64;
SEQ	SEQUENCE

CC prostrate neoplastic condition in an individual. Inhibitors of the  
CC proteins are useful for treating or preventing the progression of a  
CC prostate neoplastic condition.

SO Sequence 492 AA;

Query Match 99.2%; Score 2696; DB 21; Length 492;  
Best Local Similarity 98.8%; Pred. No. 1.2e-201;  
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 MAINSGSPPAIGRYENHGYOPENPYPAOTVPTVEVHPAOTYSPVPOVAPRYTQA 60  
DB 1 mainsgspaiispyenhygpenypaqptvtyevhpapqypqyaprvltqa 60  
OY 61 SNPVCTQPSPGTCTSTKRALCITTLTGTFELVGAALAGLTKFMSKCSNSGIEC 120  
DB 61 snpvcctqpspgtctstktkalcitltgtfelygaalaagllwtkfmskcsngiec 120  
OY 121 DSSGTCINPSNMGDGVSHCPGEGEDENRCVRLYGPNTLOYSSQKRSMPVCCODNNENY 180  
DB 121 dssgtcinpsnmgdgvshcpggedenrcvrllygpntflqyssqkrswhpvcqddneny 180  
OY 181 GRACRDMGKKNFYSSOGIVDDSGSTSEFKLNTSAGNVDTYKLYHSDACSKAVYSLR 240  
DB 181 graacrdmgyknfyssqglvddsgstsfmkltlsagnvdtlyklyhsdacsakavyslr 240  
OY 241 CIACGVNLSSRSRIYVGSALPGAMPQVSLHYQNVHVCGSITPEWITYAAHCVER 300  
DB 241 ciacgvnlssrsriyvgesalpgampqvslyhqvnhvcgsiltpewityaaahcver 300  
OY 301 PLNNPMTAFAGILRQSEMFYAGYOVERKISHPNYDSKTKNNDAIMLKLQRLTFNDL 360  
DB 301 plnnpmtafagilrqsemfyagvyoverkishpnvdsctknnidalmlklqrltfndl 360  
OY 361 VRPVCLPNEGMLQPEQLCWSGATGEEKTSEVLNAKVLLIETQCNSTRYVDNLI 420  
DB 361 vrpvcclpnegmlqpeqlcwsगतеектсевлнаквиллетқснстрывднли 420  
OY 421 TPAMICAGFLOGNVDSQCGDGPPLVTSKNNIMWLIGDTSMSGCAKARPGYGNVWF 480  
DB 421 tpmicagflognvdsqcgdspplvtsnnimlwljgdtswsgcakayrpgygnvmwf 480  
OY 481 TDWITRQMRADG 492  
DB 481 tdwityrqmkang 492

RESULT 5  
ID Y57280 standard; Protein: 492 AA.  
XX AC Y57280;  
XX DT 06-JUN-2000 (first entry)  
XX DE Ovr115 homolog protein.  
XX CS: cancer specific gene; cancer; gynecologic cancer; ovarian; breast;  
XX endometrial; uterine; lung; cytotoxic.  
XX OS Homo sapiens.  
XX PN WO200012758-A1.  
XX PD 09-MAR-2000.  
XX PF 01-SEP-1999; 99MO-US19655.  
XX PR 02-SEP-1998; 98US-0098880.  
XX PA (DIAD-) DIADEXUS LLC.  
XX

PI Salceda S, Sun Y, Reclapon H, Caferkey R;

XX WPI, 2000-256657/22.  
DR N-PSDB; Z90478.  
XX

PT Diagnosing, staging, monitoring, imaging and treating cancer especially  
PT gynecological cancers e.g. breast, ovarian cancer and lung cancer,  
PT involves measuring cancer specific gene levels in cells and body fluids

PS Disclosure; Page 52-54; 58pp; English.

CC The invention relates to detecting, diagnosing metastasis and staging  
CC cancer by measuring levels of cancer specific genes (CSG) in cells,  
CC tissues or body fluids. Their remission and progression, decreases and  
CC increases in CSG levels, is also monitored, by periodic sample analysis.  
CC The methods are useful for detecting cancers, especially gynecologic  
CC cancers which include ovarian, breast, endometrial and uterine cancer  
CC and lung cancer. Antibodies against the CSG labeled with paramagnetic  
CC ions or a radioisotope is useful for imaging cancer and when conjugated  
CC with a cytotoxic agent are useful for treating cancer. The present  
CC sequence represents a Ovr115 homolog protein, that can be used  
CC for the detection of the various cancers.

SO Sequence 492 AA;

Query Match 99.2%; Score 2696; DB 21; Length 492;  
Best Local Similarity 98.8%; Pred. No. 1.2e-201;  
Matches 486; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 MAINSGSPPAIGRYENHGYOPENPYPAOTVPTVEVHPAOTYSPVPOVAPRYTQA 60  
DB 1 mainsgspaiispyenhygpenypaqptvtyevhpapqypqyaprvltqa 60  
OY 61 SNPVCTQPSPGTCTSTKRALCITTLTGTFELVGAALAGLTKFMSKCSNSGIEC 120  
DB 61 snpvcctqpspgtctstktkalcitltgtfelygaalaagllwtkfmskcsngiec 120  
OY 121 DSSGTCINPSNMGDGVSHCPGEGEDENRCVRLYGPNTLOYSSQKRSMPVCCODNNENY 180  
DB 121 dssgtcinpsnmgdgvshcpggedenrcvrllygpntflqyssqkrswhpvcqddneny 180  
OY 181 GRACRDMGKKNFYSSOGIVDDSGSTSEFKLNTSAGNVDTYKLYHSDACSKAVYSLR 240  
DB 181 graacrdmgyknfyssqglvddsgstsfmkltlsagnvdtlyklyhsdacsakavyslr 240  
OY 241 CIACGVNLSSRSRIYVGSALPGAMPQVSLHYQNVHVCGSITPEWITYAAHCVER 300  
DB 241 ciacgvnlssrsriyvgesalpgampqvslyhqvnhvcgsiltpewityaaahcver 300  
OY 301 PLNNPMTAFAGILRQSEMFYAGYOVERKISHPNYDSKTKNNDAIMLKLQRLTFNDL 360  
DB 301 plnnpmtafagilrqsemfyagvyoverkishpnvdsctknnidalmlklqrltfndl 360  
OY 361 VRPVCLPNEGMLQPEQLCWSGATGEEKTSEVLNAKVLLIETQCNSTRYVDNLI 420  
DB 361 vrpvcclpnegmlqpeqlcwsगतеектсевлнаквиллетқснстрывднли 420  
OY 421 TPAMICAGFLOGNVDSQCGDGPPLVTSKNNIMWLIGDTSMSGCAKARPGYGNVWF 480  
DB 421 tpmicagflognvdsqcgdspplvtsnnimlwljgdtswsgcakayrpgygnvmwf 480  
OY 481 TDWITRQMRADG 492  
DB 481 tdwityrqmkang 492

RESULT 6  
ID Y81492 standard; Protein: 283 AA.  
XX Y81492;  
XX

```

FT DISULFID 788 912 INTERCHAIN (BY SIMILARITY).
FT DISULFID 826 842 BY SIMILARITY.
FT DISULFID 926 993 BY SIMILARITY.
FT DISULFID 957 972 BY SIMILARITY.
FT DISULFID 983 1011 BY SIMILARITY.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 486 486 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 762 762 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 864 864 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 903 903 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 166 192 MISSING (IN SHORT ISOFORM).
FT CONFLICT 808 808 R -> Y (IN REF. 3).
SQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CRC64;

```

```

Query Match 24.38; Score 660; DB 1; Length 1035;
Best Local Similarity 33.98; Pred. NO. 1.2e-41;
Matches 150; Conservative 67; Mismatches 180; Indels 46; Gaps 12;

```

```

OY 65 VCTPKSPSGVCTSKTKALCTLT-----GTFVGNALAGLIMKMGKSCNSG 117
DB 611 VYTPGPGVNVDFSTNMATVLFITDNMLAKGFKANFTTGGLG-----IPPCKEDN 663
OY 118 IECDSSTGTCINPSCMGVSHCGEGEDENRCVLYG-----PNFLOYSSQKRSWH 169
DB 664 FQC-KDGEICIPVLNLCDFPCKDGSDEAHCHVRLFNCTDSSGLVQFRIQ-----SIWH 716
OY 170 PVCDODMNNENGRACADMGKNNFYSSGGLVDSGSSFPKMLTMSAGNDYIKKLHSD 229
DB 717 VACENMTTQISDDVLCGLGTS--NSSVPTFSTGGGPPYVNLTAPEGSLI---LPPSQ 771
OY 230 ACSKAYAVSLRC--IAGVNLNSROS-RIYGESALPGAMPVOVSLHYOVHVGCGSII 286
DB 772 QCLEDSLILLCQNKYSCGKLVTOEVSFKTYGSDSREBAMPVVALYFDQOYCGASLV 831
OY 287 TPFWITAAHCVEKPLNPMHMTAFAGILRSFMYGAGYVE-----KYISHPNYSKT 341
DB 832 SRDLVGAHCVGRNNEPSKMKAVLGLHMSNL---TSPQIEIRLDIQIYINHYNNRR 888
OY 342 KNNDIAMLKLOKPLTFNDLVKPVCLPNPMMLOPEQOLCWSIGWATEEKGKTSYVLAIAK 401
DB 889 KNNDIAMHLEMKVNYDYDIPICLPENQVFPGRICISINGALITYOGSTADVLQESAD 948
OY 402 VLLIETORCSRNYDYDILIPAMICAGFLOGNNDSCGDSGGLVTSKNNNIMWMLIGDSW 461
DB 949 VPLLNSNECCQOMPEYN-ITENMYCAGEAGVDSGCGPLMCCNNMMLLAGVTSF 1007
OY 462 GSGCAKAYRPGVYGNVWVFTDMI 484
DB 1008 GYOCALPRNRPVYARVPRFTEMI 1030

```

```

RESULT 5
ENTR_HUMAN STANDARD; PRT; 1019 AA.
AC P8073;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)

```

```

DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
GN PRSS7 OR ENTK.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Duodenum;
RX MEDLINE=95234679; PubMed=7718557;
RA Kilmoto Y., Velle R.A., Donis-Keller H., Sadler J.E.;
RT "cDNA sequence and chromosomal localization of human enterokinasase,
the proteolytic activator of trypsinogen.";
RL Biochemistry 34:4562-4568(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Holzinger A., Buck C., Maier E.M., Meyerhofer P.U., Roscher A.A.,
RA Sadler J.E., Hadorn H.B.;
RT "Genomic organization of the human enteropeptidase.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Soeda E.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Onki M., Takagi T., Sakaki Y., Tauden S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Shlenthall A., Sasaki T., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Degen E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [4]
RP SEQUENCE OF 749-1019 FROM N.A.
RC TISSUE=Duodenum;
RX MEDLINE=94329561; PubMed=8052624;
RA Kilmoto Y., Yuan X., Wu O., McCourt D.W., Sadler J.E.;
RT "Enterokinasase, the initiator of intestinal digestion, is a mosaic
protease composed of a distinctive assortment of domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS--1-ILE-7 BOND IN
TRYPSINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- DISEASE: DEFECTS IN PRSS7 CAUSE LIFE-THREATENING INTESTINAL
MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 WAM DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```



Query Match 19.7%; Score 534; DB 2; Length 356;  
 Best Local Similarity 35.4%; Pred. No. 4.4e-41;  
 Matches 120; Conservative 53; Mismatches 132; Indels 34; Gaps 11;

169 HPCVGDMMNENYRAACRDMYKNNFYSSGIVDDSGSTFEMKLN---SAGVNDIYKKL 225  
 25 HHVCAGQWEIISQLCKQKGLCEPSVTKL-IOEQEKPRMLTLHSHWESLNTTTHLELL 83  
 226 YHSDACSKAVVSLRCLIA--CGVNLNNSRROSRIYVGSALPGAMPQVSLHVO--NVHVC 282  
 84 VNGQSESRKISILLCTKQDCGRPARBMKRLTGGRTSRGMPMPCSLQSPSHGICG 143  
 283 GSITPEWITAAHCV-----KPLNPMHTAPAGILROSEMFYAGYOEKY 331  
 144 CVLIARKWVLTVAHCFEGRENAVXKRVGLINNLDHPVSF---MOTRF-----VKTI 192  
 332 ISHPNDSKTKNNDLAMKOKPLTFNDLYKPVCLPMPGMLPEOLCMTISGKATEK 391  
 193 ILHPRYSRAVVDYDISVELSEDISETGYRVPCLPPEOMLEPDYCYITGNG--HMG 250  
 392 KTSSEVNAKAVLLIETORCNSRYVNLITPAMICAGFLOGNVDSGSGGPLYTSK-N 450  
 251 KMFKLOEGSEVRISLEHQS-YFDKMTITRMICAGYEGSTVDSGSGGPLYVEKPC 309  
 451 NIMWLIGTSMGSGC-AKAYRPGVYGNVFTMYIROM 488  
 310 GRWTLFGLTSMGSGVCSKVLGPGVYSNVSFVEMIKROI 348

## RESULT 10

US-09-027-337-3  
 ; Sequence 3, Application US/09027337B  
 ; Patent No. 5972616  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Brien, Timothy J.  
 ; APPLICANT: Tanioto, Hiroto  
 ; TITLE OF INVENTION: TABG-15: An Extracellular Serine Protease Overexpressed in  
 ; FILE REFERENCE: D6064  
 ; CURRENT APPLICATION NUMBER: US/09/027,337B  
 ; CURRENT FILING DATE: 1998-02-20  
 ; NUMBER OF SEQ ID NOS: 13  
 ; SEQ ID NO 3  
 ; LENGTH: 256  
 ; TYPE: PRP  
 ; ORGANISM: Unknown  
 ; FEATURE:  
 ; OTHER INFORMATION: Serine protease catalytic domain of hepsin (Heps)  
 ; OTHER INFORMATION: homologous to similar domain in TABG-15  
 US-09-027-337-3

Query Match 19.5%; Score 528.5; DB 2; Length 256;  
 Best Local Similarity 41.2%; Pred. No. 8.9e-41;  
 Matches 101; Conservative 41; Mismatches 92; Indels 11; Gaps 3;

255 RIVGESALPGAMPQVSLHVOVHVCSSITPEWITAAHCEKPLNPMHTAFAGI 314  
 1 RIVGGRRTSLGRMPQVSLHYDGAHLGSGILSGDWLTAAHCEPENRVLSTRMRYAGA 60  
 315 LRSEMFYAGYOEKYISHPN-----DSKTKNDIALMKLOKPLTFNDLYKPVCLPN 368  
 61 VAQASP-HGLQLOVQAVVYGGYLPFRDPNSENNDIALVHLSSPLPLEYIOPVCLPA 119  
 369 PGWMLPEOLCMTISGKATEKTSSEVNAKAVLLIETORCNSRYVNLITPAMICAG 428  
 120 AGALVDGKICTYTGKNTQYTGQAGVLOEAVPIISNVYCNAGADPYGQIRPKPCAG 179  
 429 FLOGNVDSGSGGPLY---TSKNIMWLIGDTSMGSCAKAYRPGVYGNVFTMYIRQ 484  
 180 YPBGIDACGSGGSPVCEDSISRPRMRLGIVSGTCALAKRGVYTKVSDPREMI 239  
 485 YROMR 489

DB 240 FOAIR 244

## RESULT 11

US-08-508-448C-25  
 ; Sequence 25, Application US/08508448C  
 ; Patent No. 5804410  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kazuyoshi YAMAKA et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING  
 ; TITLE OR INVENTION: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Wenderoth, Lind & Ponack  
 ; STREET: 805 Fifteenth Street, N.W., #700  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: U.S.A.  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: Wordperfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/508,448C  
 ; FILING DATE: July 28, 1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Warren M. Cheek, Jr.  
 ; REGISTRATION NUMBER: 33,367  
 ; REFERENCE/DOCKET NUMBER:  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-371-8850  
 ; TELEFAX:  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 25:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 418 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 US-08-508-448C-25

Query Match 19.4%; Score 527.5; DB 1; Length 418;  
 Best Local Similarity 39.8%; Pred. No. 2.2e-40;  
 Matches 111; Conservative 43; Mismatches 116; Indels 9; Gaps 6;

212 LNTSANDVDIKKLYHSDACSKAVVSLRCLICGY--NINSSROSRIYVGSALPGAMP 269  
 143 LNSN-GNLEINPST-BITSITDQAAANMLINBCGAGPDILITSEQIILLGTAEBSMFW 200  
 270 QVSLHVOVHVCSSITPEWITAAHCEKPLNPMHTAFAGILROSEMFYAGYOE 329  
 201 QVSLRLNNAHHCGLINNMWILTAHCF-RSNSNPRDIATSGI---STTPFKLMBRV 256  
 330 KYISHPNYSKTKNNDLAMKLOKPLTFNDLYKPVCLPMPGMLPEOLCMTISGKATE 389  
 257 NILIHNNYSATHEINDIALVRLNSVTFKDIHSVCLPATONIPGSAVYTGWGAOEY 316  
 390 KGTSEVNAKAVLLIETORCNSRYVNLITPAMICAGFLOGNVDSGSGGPLYTS 448  
 317 AGTVEBELRGQVRIISNDVCNAPHYSYNGAILSGMLCAGVPAGVDACGSGGPLYVED 376  
 449 KNIMWLIGDTSMGSCAKAYRPGVYGNVFTMYIRQ 487  
 377 SRLMFTIVGIVSGDGLPDRGVYTRVATYADWIRQ 415



NCBI\_Taxid-10090;  
 (1)  
 SEQUENCE FROM N.A.  
 STRAIN-C57BL/6; TISSUE-Duodenum;  
 MEDLINE-98147142; PubMed-9486188;  
 Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;  
 Structure of murine enterokinase (enteropeptidase) and expression in  
 small intestine during development.";  
 Am. J. Physiol. 274:G342-G349(1998).  
 - FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC  
 PROTEOLYTIC PHENOMENES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE  
 A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN  
 TURN ACTIVATES OTHER PHENOMENES INCLUDING CHYMOTRYPSINOGEN,  
 PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).  
 - CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN  
 TRYPSINOGEN.  
 - SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A  
 MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY  
 SIMILARITY).  
 - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).  
 - PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS  
 CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).  
 - SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.  
 - SIMILARITY: CONTAINS 2 SCRB DOMAINS.  
 - SIMILARITY: CONTAINS 1 MAM DOMAIN.  
 - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPSIN FAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: U73378; AAB3317.1; -  
 HSSP: P00763; IDPO.  
 MEROPS: S01.156; Prs57.  
 MGD: MGI1197533; Prs57.  
 InterPro: IPR000082; -  
 InterPro: IPR000859; -  
 InterPro: IPR000988; -  
 InterPro: IPR001190; -  
 InterPro: IPR001254; -  
 InterPro: IPR001314; -  
 InterPro: IPR002172; -  
 Pfam: PF00431; CUB; 2.  
 Pfam: PF00629; MAM; 1.  
 Pfam: PF01390; SEA; 1.  
 Pfam: PF00530; SRCR; 1.  
 Pfam: PF00057; LDL\_recept\_a; 2.  
 Pfam: PF00089; LTPsin; 1.  
 PRINTS: PR00722; CHYMOTRYPSIN.  
 PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 PROSITE: PS00135; TRYPSIN\_SER; 1.  
 PROSITE: PS0180; CUB; 2.  
 PROSITE: PS00740; MAM\_2; 1.  
 PROSITE: PS0180; MAM\_1; 1.  
 PROSITE: PS01209; LDLRA\_1; 2.  
 PROSITE: PS50068; LDLRA\_2; 2.  
 Signal-anchor: Glycoprotein; Myristate; Hydroxylase;  
 Kunitz protease; zymogen; Transmembrane; Repeat.  
 Kunitz protease: zymogen; Transmembrane; Repeat.  
 Non-catalytic chain (heavy chain).  
 CATALYTIC CHAIN (LIGHT CHAIN).  
 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 LDL-RECEPTOR CLASS A 1.  
 CUB.  
 MAM.  
 CUB.  
 LDL-RECEPTOR CLASS A 2.  
 SRCR.

FT	ACT_SITE	874	874	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	925	925	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	1021	1021	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	LIPID	2	2	MYRISTATE (POTENTIAL).
FT	DISULFID	229	242	BY SIMILARITY.
FT	DISULFID	236	255	BY SIMILARITY.
FT	DISULFID	249	266	BY SIMILARITY.
FT	DISULFID	688	700	BY SIMILARITY.
FT	DISULFID	695	713	BY SIMILARITY.
FT	DISULFID	707	722	BY SIMILARITY.
FT	DISULFID	817	945	INTERCHAIN (BY SIMILARITY).
FT	DISULFID	859	875	BY SIMILARITY.
FT	DISULFID	959	1027	BY SIMILARITY.
FT	DISULFID	991	1006	BY SIMILARITY.
FT	DISULFID	1017	1045	BY SIMILARITY.
FT	CARBOHYD	147	147	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	197	197	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	212	212	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	373	373	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	380	380	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	433	433	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	515	515	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	579	579	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	675	675	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	727	727	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	751	751	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	770	770	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	791	791	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	897	897	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	936	936	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	999	999	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	1069 AA;	118735 MW;	BE2549E463743C3D CMC64;

Query Match 23.5%; Score 638; DB 1; Length 1069;  
 Best Local Similarity 30.5%; Pred. No. 5,3e-40;  
 Matches 164; Conservative 92; Mismatches 196; Indels 86; Gaps 21;

QY	18 HGQPEPNRYPAQTPVPT	-----YVEV-HPAQYPPSPVQYAPVYLQAS-----	61
DB	544 NGICSGSPY-EPFLVPTPPPELPTDCGPFLEPNSTPSS--PNPDPKYPNQASCIWN	600	
QY	62 -----NPV-----	CTQPKSPSGVCSKTKAL	85
DB	601 LNAQGRKNIOLHQEFLEININDVEVVDGGEFDSLLAVYGP-GVVKDLSTTNKATV	659	
QY	86 CITLLGFLVYG--ALALAGLTKFNG--SKCSNSGIECDSSGTCINPNSMCDGVSHPG	141	
DB	660 IFTTMMETRRKGFKAANTSQY--YLGIPRCODDEQC-KDGNCTPLGLNCLDSYPRCRD	715	
QY	142 GEDENRCVRLY---GNFTLOYYSQSKSMHPYVQDDMNENYGRAACRDMGKNNFYSS	197	
DB	716 GSDASCVRFNGTRNSNGVLQF--NIHSIMHICAEIMWTQISNEVCHLGLGS--ANS	771	
QY	198 QGTYDDSGSTSFMLKNTSAGVNDLYKLIHSDACSSAAVSLRC--IAGCVNLSSNOS-	254	
DB	772 SMPISSTGGEFVAVNAPNGSLI--LTPSLQCSQSLILLCNHRKSCGEKAYTVSP	828	
QY	255 RIVGESALPGAMPQVYSLAQNVH---VCGGSITPENIVTAACVERPLNPMHTA	310	
DB	829 KIVGSDAQAGAMPVYALKLRDSTRLCGASVSDMLYSAAHCVYRNINDPTMTA	888	
QY	311 FAGILROSFMEYGAQY--VEKVISHPNYSKTKNNDIALMKLQKPLTFNDLYKVCPLN	368	
DB	889 VLGHMOSNLITSPVVRVAVQDIYNPHYDRKRVNDIAMHLEFKVNYDYIOLPILPE	948	
QY	369 PGAMLOEOLCISGNKATE-EKKTSEVYLNAAKVLILFQRCRSRVYDNLTPAMICA	427	
DB	949 ENQIFIGRTCSIGAWGYDKINASTVDYKADVPILSNKCCQOQAPERYN-ITSEWICA	1007	
QY	428 GFLGNDVSCGDSGGFLVYTSKNNIMLIGDTSGWGCACAAAYRPGVYGNVAFPTDWTY	485	
DB	1008 GYEGGIDSCGDSGGFLGMOENNRMTFLVGTSTGVCACALPNHGVYVVRYSQFLEWH	1065	

Db 772 CFEBSLLILLCNNAHSCGKQVAVSVSKYVGNDSREGANPWTVALYYNGLLGASIVS 831  
 QY 288 PEMIVTAHCVKPEKLPNPMHTAFAGILROSFMYGAGYQ-----VEKVISHPYDSKTK 342  
 Db 832 RDMVSAHACHVYGNLLEPSSKKAILGLHMTSNTL---TSPQVTLIDELIVINPHYNRKK 888  
 QY 343 NNDIALMLKLPFLFNDLVKPVCLPNPGMMLOPQOLWISWGATEEKGTSEVLNAAKV 402  
 Db 889 DSDIAMHLEKRYVYTDYIOPICLPEENOVFPGRICSIAGMGVITYOGSPADILQEDAV 948  
 QY 403 LLITQRCNSYVVDNLITPAMICAGFLOGVNDSCOGSGGSLTYSKNIMVLLIGDTSMG 462  
 Db 949 PLTSENCCQOQMPRYN-ITENMMCAGTBEGSIDSCQDGGSLMCLNNKLLAGVTSFG 1007  
 QY 463 SGCAKAYRPGYGVNMYFTDWI 484  
 Db 1008 YQCALPMRPYVYARVPKTEWI 1029

RESULT 8  
 KAL\_HUMAN STANDARD; PRT; 638 AA.  
 AC P03952;  
 DT 23-OCT-1986 (Rel. 02, Created)  
 DT 23-OCT-1986 (Rel. 02, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)  
 DE (KININOGENIN) (FLETCHER FACTOR).  
 GN KIK3.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86243359; PubMed=3521732;  
 RA Chung D.W., Fujikawa K., McMullen B.A., Davie E.W.;  
 RT "Human plasma prekallikrein, a zymogen to a serine protease that  
 RT contains four tandem repeats.";  
 RL Biochemistry 25:2410-2417(1986).  
 RN [2]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=91152016; PubMed=1998666;  
 RA McMullen B.A., Fujikawa K., Davie E.W.;  
 RT "Location of the disulfide bonds in human plasma prekallikrein: the  
 RT presence of four novel apple domains in the amino-terminal portion of  
 RT the molecule.";  
 RL Biochemistry 30:2050-2056(1991).  
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT  
 CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING  
 CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM  
 CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN  
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.  
 CC -1- SUBUNIT: THE ZMOGEN IS ACTIVATED BY FACTOR XII, WHICH CLEAVES  
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,  
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE  
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.  
 CC -1- DISEASE: DEFECTS IN KIK3 ARE THE CAUSE OF FLETCHER FACTOR  
 CC DEFICIENCY, A BLOOD COAGULATION DEFECT.  
 CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: M31143; AAA60153.1; -  
 CC PIR: A00921; KOHUP.

DR PIR: A37939; A37939.  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.212; -.  
 DR MIM; 229000; -.  
 DR InterPro: IPR000177; -.  
 DR InterPro: IPR001254; -.  
 DR InterPro: IPR001314; -.  
 DR InterPro: IPR003014; -.  
 DR Pfam; PF00024; PAN; 4.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00005; APPLIEDOMAIN.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR PROSITE; PS00495; APPLE; 4.  
 DR KMW Hydrolysis: Serine protease; Glycoprotein; Plasma; Zymogen; Signal;  
 DR KMW Fibrinolysis: Blood coagulation; Inflammatory response; Liver;  
 KW Repeat.  
 FT SIGNAL 1 19  
 FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.  
 FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.  
 FT DOMAIN 20 105 APPLE 1.  
 FT DOMAIN 110 195 APPLE 2.  
 FT DOMAIN 200 285 APPLE 3.  
 FT DOMAIN 291 376 APPLE 4.  
 FT DOMAIN 389 621 CATALYTIC.  
 FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .).  
 FT ACT\_SITE 434 434 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 483 483 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 578 578 CHARGE RELAY SYSTEM.  
 FT DISULFID 21 104  
 FT DISULFID 47 77  
 FT DISULFID 51 57  
 FT DISULFID 111 194  
 FT DISULFID 137 166  
 FT DISULFID 141 147  
 FT DISULFID 201 284  
 FT DISULFID 227 256  
 FT DISULFID 231 237  
 FT DISULFID 292 375  
 FT DISULFID 318 347  
 FT DISULFID 322 328  
 FT DISULFID 340 345  
 FT DISULFID 383 503  
 FT DISULFID 419 435  
 FT DISULFID 517 584  
 FT DISULFID 548 563  
 FT DISULFID 574 602  
 SO SEQUENCE 638 AA; 71369 MW; E62P9C1053838FB4 CRC64;

Query Match 21.0%; Score 569.5; DB 1; Length 638;  
 Best Local Similarity 32.8%; Pred. No. 37e-35;  
 Matches 155; Conservative 64; Mismatches 160; Indels 93; Gaps 23;

QY 55 RVLTQASNPVYCTQPKPSGVCISKTRKALCTLTGTPLVGAALAGLLMKPMGS--- 111  
 Db 218 RVLT--DPAFCR-----TICTYHNP--CLFTEYTT-----NWKLTESGRNV 235  
 QY 112 ---RCSNNGICRDS-----SG-----TCINP-----SNMCDVSHCPGGEDENRVRL 151  
 Db 256 CLKLTSSGSPSSSTPQENTISGSLTCKRTLPKCHSKRIYVDV--GGEELN----- 308  
 QY 152 YGPNFLQVYSSQKSWHPYVCOQDWMENYGRACRDMGYKNNEYSSGIVDSSGTS--- 208  
 Db 309 --YVFVKGV-----NVCQE-----TCTKM-IRCOFTFYSLPDEDCKEERCKC 347  
 QY 209 FMKLNTAGNVDYKIKLYHSDACSSKAVVSLRCLACGVN--LNSRSRSRTVGSGSLRPGA 266

DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE SUPPRESSOR OF TUMORIGENICITY 14 (EC 3.4.21.-) (MATRIPTASE) (MEMBRANE-  
 DE TYPE SERINE PROTEASE 1) (MT-SP1).  
 GN S714 OR PRS514 OR SNC19.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=99303581; Pubmed=10373424;  
 RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;  
 RA "Molecular cloning of cDNA for matriptase, a matrix-degrading serine  
 RA protease with trypsin-like activity."  
 RT J. Biol. Chem. 274:18231-18236(1999).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA Takeuchi T., Shuman M.A., Craik C.S.;  
 RA "Reverse biochemistry: Use of macromolecular protease inhibitors to  
 RT dissect complex biological processes and identify a membrane-type  
 RT serine protease in epithelial cancer and normal tissue."  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).  
 RL [3]  
 RP CHARACTERIZATION.  
 RP TISSUE=Milk;  
 RX Pubmed=10373425;  
 RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;  
 RA "Purification and characterization of a complex containing matriptase  
 RT and a Kunitz-type serine protease inhibitor from human milk."  
 RL J. Biol. Chem. 274:18237-18242(1999).  
 CC -1- FUNCTION: DEGRADERS EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE  
 CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE  
 CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG  
 CC OR LYS AS THE P1 SITE.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).  
 CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL: AF118224; AAD42765.2;  
 DR EMBL: AF133086; AAF00109.1;  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.302;  
 DR InterPro: IPR000859;  
 DR InterPro: IPR001254;  
 DR InterPro: IPR001314;  
 DR InterPro: IPR002172;  
 DR Pfam: PF00057; 1dl\_recept\_A; 4.  
 DR Pfam: PF00089; trypsin\_1.  
 DR Pfam: PF00431; CUB.2.  
 DR PRINTS: PRO0261; LDLRECEPTOR.  
 DR PRINTS: PRO0722; CHYMOTRYPSIN.  
 DR PROSITE: PS00134; TRYPSIN\_HIS.1.  
 DR PROSITE: PS00135; TRYPSIN\_SER.1.  
 DR PROSITE: PS01180; CUB.2.  
 DR PROSITE: PS01209; LDLRA\_1; 2.  
 DR PROSITE: PS50068; LDLRA\_2; 4.  
 DR Signal-anchor: Glycoprotein; Hydrolase; Serine protease;  
 KM Transmembrane; Repeat.  
 FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).  
 FT TRANSLEM 56 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT TRANSLEM 56 76 (POTENTIAL).  
 FT DOMAIN 77 855 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 214 334 CUB 1.  
 FT DOMAIN 340 447 LDL-RECEPTOR CLASS A 1.  
 FT DOMAIN 452 487 LDL-RECEPTOR CLASS A 2.  
 FT DOMAIN 487 524 LDL-RECEPTOR CLASS A 3.  
 FT DOMAIN 524 560 LDL-RECEPTOR CLASS A 4.  
 FT DOMAIN 566 603 CATALYTIC.  
 FT DOMAIN 614 851 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 656 656 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 711 711 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 711 711 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CARBOHYD 109 109 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 302 302 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 485 485 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 772 772 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SO SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;  
 Query Match 20.6%; Score 558.5; DB 1; Length 855;  
 Best Local Similarity 33.9%; Pred. No. 3.4e-34;  
 Matches 131; Conservative 57; Mismatches 129; Indels 69; Gaps 14;  
 QY 110 GSKGNSGIECDSSGTCINPMDGVSHCPGDEHRCVRLYGNFLQYSSQKSMH 169  
 Db 522 GCSCPAQTFRC-SNGKLSKSGQCNKGDGSDSASCPV-----NVV----- 565  
 QY 170 PVCODDMNNTGRACSDMGYKNNFYSSQGITVDGSGTSEKMLTSGANDYKLYHSD 229  
 Db 566 -----TCT-----KHYRCLNG-----LTSKGNPECDKEDCSD 595  
 QY 230 ACSKAYSLRCLGCVNLN-SQSRIVGESALPQAMPQVSLH-VQNVHGGGSIIT 287  
 Db 596 GSDEK-----DCDGLRFRGTARVAGGTDADEEPMQVSLHLDGCHIGASLIS 647  
 QY 288 PEWYTAHCVKEP-----LNNPMHTAFAGILRSFMEYAGAYO---VEKISHPNYDSK 340  
 Db 648 PNLVSAHACHIDDRGRFSDPTQMTAFGLHDS-QRSAPGVERLKRISHPFNDF 706  
 QY 341 TKNDIALMKLQKPLFNDLVKPYCLPQPMGLPQPOLCWTISGATBEKGISEVLNAA 400  
 Db 707 TFDYDIALLELEKPAEYSSMVRPCLDPASHVFPAGKAIVTQGHGHOYGTGALLILQK 766  
 QY 401 KVLIEFQRCNSRVYDNLIRPAMICAGFLOGNVDSCGDSGGLP-VTSNNIIMWLGID 458  
 Db 767 EIRYNOTTCEN--LTPQITPRMVCGLFSGVYDSCGDSGGLSVYEDGRI-FOAGV 823  
 QY 459 TSMGSCAKAYRPGVYGVNVTDM 484  
 Db 824 VSMGDGCAQRNKPQVYTRRLPFRD 849  
 RESULT 11  
 KAL\_RAT STANDARD; PRT; 638 AA.  
 ID KAL\_RAT  
 AC P14272;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)  
 DE (KININOGENIN) (FLETCHER FACTOR).  
 GN PK.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=91129236; Pubmed=1993180.  
 RA Beaubien G., Rosinski-Chupin I., Mattei M.-G., Mikiy M., Chretien M.,  
 RA Seidah N.G.;  
 RA "Gene structure and chromosomal localization of plasma kallikrein."  
 RT Biochemistry 30:1628-1635(1991).  
 RL [2]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN INTESTINE, KIDNEY, LUNG,  
 CC AND THYMUS. NOT EXPRESSED IN SKELETAL MUSCLE, LIVER, HEART.  
 CC TESTIS, AND BRAIN.  
 CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPsin FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF042822; MAD02230.1; -  
 DR MGD: MGI:133881; St14.  
 DR MEROPS: S01.014; -  
 DR InterPro: IPR000859; -  
 DR InterPro: IPR001254; -  
 DR InterPro: IPR001314; -  
 DR InterPro: IPR002172; -  
 DR Pfam: PF00431; CUB; 2.  
 DR Pfam: PF00057; Ldl\_recept\_a; 4.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00261; LDLRECEPTOR.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PROSITE: PS01209; LDLRA\_1; 2.  
 DR PROSITE: PS50068; LDLRA\_2; 4.  
 DR PROSITE: PS01180; CUB; 2.  
 DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE: PS00135; TRYPsin\_SER; 1.  
 DR Signal-anchor; Glycoprotein; Hydrolase; Serine protease;  
 DR Transmembrane; Repeat.  
 FT DOMAIN 1 55  
 FT TRANSMEM 56 76  
 FT 77 902  
 FT DOMAIN 214 331  
 FT 340 444  
 FT DOMAIN 451 488  
 FT 489 522  
 FT DOMAIN 523 561  
 FT 565 604  
 FT DOMAIN 614 851  
 FT 851 851  
 FT ACT\_SITE 656 656  
 FT ACT\_SITE 711 711  
 FT ACT\_SITE 805 805  
 FT CARBOHYD 107 107  
 FT CARBOHYD 302 302  
 FT CARBOHYD 365 365  
 FT CARBOHYD 421 421  
 FT CARBOHYD 489 489  
 FT CARBOHYD 772 772  
 FT SEQUENCE 902 AA; 99645 MW; 2A3ED02AFA69C04E CRC64;  
 Query Match 20.2%; Score 548.5; DB 1; Length 902;  
 Best Local Similarity 32.7%; Pred. No. 2e-33;  
 Matches 123; Conservative 66; Mismatches 118; Indels 69; Gaps 13;

DB 594 SDGSDEK-----NC-DGGLR-SFTKQARYGCTNADSGEMPVOVSLHALGCHLCSASLI 646  
 QY 287 TPENIVTAACVCKEPLNNPMH---WTAFAGLIKQSFMTYAGYQ---VERVISHPNYDS 339  
 DB 647 SPDLVSAACHFODDKFKKSDYMTMTAFGLIDQS-KRSASQVQEKLRITTHPSFND 705  
 QY 340 KTKNDNALKLLQKPLTFNDLVKPVCLPNPGMMQLPPQQLMISGMGTEKGTSEVLNA 399  
 DB 706 FTFDYDIALLEKSVSEYTVVRPCLPDATVHPACKAIWGTGHTKGGTGALTLQK 765  
 QY 400 AKVLIETORNSRYVDNLITPAMICAGFLQGNVDSGQDSGGLVTS-KNNIMWLGID 458  
 DB 766 GEIVINOTCED--LMPQDITPMKCVGLSGYVDSGQDSGLPSAKEDGMPAGV 823  
 QY 459 TSMGSGCAKAVRPGVY 474  
 DB 824 VSMGEGCAQRNKPQVY 839  
 RESULT 13  
 ID KAL\_MOUSE STANDARD; PRT; 638 AA.  
 AC P26262;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)  
 DE (KININOGENIN) (FLETCHER FACTOR).  
 GN KLK3 OR PK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RC SPRAIN-BALB/C; TISSUE=Liver;  
 RX MEDLINE=91090844; PubMed=2264928;  
 RA Seidman N.G., Sawyer N., Hamelin J., Mion P., Beaudien G.,  
 RA Brechpapa L., Rochmont J., Mbikay M., Chretien M.;  
 RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,  
 RT and comparison of protein and mRNA levels among species.";  
 RL DNA Cell Biol. 9:737-748(1990).  
 CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT  
 CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING  
 CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM  
 CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN  
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.  
 CC -1- SUBUNIT: THE ZMOGEN IS ACTIVATED BY FACTOR XIIa, WHICH CLEAVES  
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,  
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE  
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.  
 CC -1- SIMILARITY: CONTAINS 4 PEPTIDASE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPsin FAMILY. PLASMA KALLIKREIN SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M58588; AAA63393.1; -  
 DR PIR: A36587; KOMSPL.  
 DR HSSP: P00750; IRTF.  
 DR MEROPS: S01.212; -  
 DR MGD: MGI:102849; KIK3.  
 DR InterPro: IPR000177; -  
 DR InterPro: IPR001254; -  
 DR InterPro: IPR001314; -  
 DR InterPro: IPR003014; -  
 DR Pfam: PF00024; PAN; 4.

```

OY 137 SHCGGDEKRCVRLYPNFILOYSSQRKSMHPVCODMNNENYGRAACRDMGKNNFYS 196
DB 41 ILLOSDEPLYQVOLSGDSRLAVLADTGTETWLLCSSRSNARAGLCCEBMGLRLALAH 100
OY 197 SOGIVDSDSGTSEFKMLNTSA-----GNVDIYKKLYHSDA---CSSKAVYSLRCIACGVN 247
DB 101 SELDVRNAGAN-----GTSGFFCVDEGGLPLAQLRLLVISVCDPRPRFLATQDDG-- 153
OY 248 LNSRSRIGVGSALPGAMPQVSLHVQNVHCGSGIIPENIVTAHCEKPLNPMH 307
DB 154 RRLKPVDRIVGGOSSSLGRMPQWYSLRYDGTHTLCCGSLLSGDWVLTAAHCFERNRLSR 213
OY 308 WTAFAAGLRBSFMYGAGVYKATISPNY-----DSKKNNDIAKMLQKPLTFNDLV 361
DB 214 WRFFAGAVARTSP-HAVALGQVAVIYHGGLPERDPTIDENSNDIAVLHSSSLPLEYEY 272
OY 362 KPVCLPAPGMMLOPEOLCMTISGMATEKTESEVLNAAKLLIETORCNSRYVDNLIT 421
DB 273 QPVCLPAGQALDVGKCYTGTGMCNTQFYGOQANVLOEAPRITISNEVCSPDPYQNDIK 332
OY 422 PAMICAGFLOGMNDSCOGDSGGPLY---TSKNIMWLIGDTSNGSCANARPGYGVN 477
DB 333 PKMFCAGYPEGIDACGDSGPFVCEDSISGTSRMRLCGIVSGTGALARKPGYTKV 392
OY 478 MWFTDWIYKRM 489
DB 393 TDFREWIFKAIR 404

RESULT 15
FAIL_HUMAN STANDARD; PRT; 625 AA.
ID P03951;
AC 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COAGULATION FACTOR XI PRECURSOR (EC 3.4.21.27) (PLASMA THROMBOPLASTIN
DE ANTECEDENT) (PTA).
GN F11.
OS Homo sapiens (Human).
OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP MEDLINE-6243360; PubMed-2836155;
RA Fujikawa K., Chung D.W., Hendrickson L.E., Davie E.W.;
RT "Amino acid sequence of human factor XI, a blood coagulation factor
RT with four tandem repeats that are highly homologous with plasma
RT prekallikrein."
RL Biochemistry 25:2417-2424(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-68107663; PubMed-2827746;
RA Aakai R., Davie E.W., Chung D.W.;
RT "Organization of the gene for human factor XI."
RL Biochemistry 26:7221-7228(1987).
RN [3]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE-91152017; PubMed-1998667;
RA McMullen B.A., Fujikawa K., Davie E.W.;
RT "Location of the disulfide bonds in human coagulation factor XI: the
RT presence of tandem apple domains."
RL Biochemistry 30:2056-2060(1991).
RN [4]
RP VARIANT LEU-301.
RX MEDLINE-90046636; PubMed-2813350;
RA Aakai R., Chung D.W., Ratnoif O.D., Davie E.W.;
RT "Factor XI (plasma thromboplastin antecedent) deficiency in Ashkenazi
RT Jews is a bleeding disorder that can result from three types of point
RT mutations."
RL Proc. Natl. Acad. Sci. U.S.A. 86:7667-7671(1989).

```

```

RN [5]
RP VARIANT LEU-301.
RX MEDLINE-92190478; PubMed-1547342;
RA Meijers J.C., Davie E.W., Chung D.W.;
RT "Expression of human blood coagulation factor XI: characterization of
RT the defect in factor XI type III deficiency."
RL Blood 79:1435-1440(1992).
CC -1- FUNCTION: FACTOR XI TRIGGERS THE MIDDLE PHASE OF THE INTRINSIC
CC PATHWAY OF BLOOD COAGULATION BY ACTIVATING FACTOR IX.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF ARG-1-ALA AND ARG-1-VAL
CC BONDS IN FACTOR IX TO FORM FACTOR IXA.
CC -1- SUBUNIT: HOMODIMER; LINKED BY A DISULFIDE BOND. AFTER ACTIVATION
CC THE HEAVY AND LIGHT CHAINS ARE ALSO LINKED BY A DISULFIDE BOND.
CC -1- PTM: ACTIVATED BY FACTOR XIIA (OR XII), WHICH CLEAVES EACH
CC POLYPEPTIDE AFTER ARG-387 INTO THE LIGHT CHAIN, WHICH CONTAINS THE
CC ACTIVE SITE, AND THE HEAVY CHAIN, WHICH ASSOCIATES WITH HIGH
CC MOLECULAR WEIGHT (HMW) KININOGEN.
CC -1- DISEASE: DEFECTS IN F11 ARE A CAUSE OF A BLOOD COAGULATION
CC ABNORMALITY (ROSENTHAL SYNDROME) OCCURRING IN HIGH FREQUENCY IN
CC ASHKENAZI JEWS.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13142; AAA52487.1; -
DR EMBL; M20218; AAA51985.1; -
DR EMBL; M18296; AAA51985.1; JOINED.
DR EMBL; M21184; AAA51985.1; JOINED.
DR EMBL; M18298; AAA51985.1; JOINED.
DR EMBL; M18299; AAA51985.1; JOINED.
DR EMBL; M18300; AAA51985.1; JOINED.
DR EMBL; M18301; AAA51985.1; JOINED.
DR EMBL; M18302; AAA51985.1; JOINED.
DR EMBL; M18303; AAA51985.1; JOINED.
DR EMBL; M18304; AAA51985.1; JOINED.
DR EMBL; M19417; AAA51985.1; JOINED.
DR EMBL; M20217; AAA51985.1; JOINED.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.213; -
DR MIM; 264800; -
DR InterPro; IPR000177; -
DR InterPro; IPR001254; -
DR InterPro; IPR001314; -
DR InterPro; IPR003014; -
DR Pfam; PR00024; PAN; 4.
DR Pfam; PR00089; trypsin; 1.
DR PRINTS; PR00005; APPLEDOMAIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS00495; APPLE; 4.
KW Hydrolyase; Serine protease; Glycoprotein; Plasma; Blood coagulation;
KW Duplication; Signal; Disease mutation.
FT SIGNAL 1
FT CHAIN 19 387
FT CHAIN 388 625
FT DOMAIN 19 104
FT DOMAIN 109 194
FT DOMAIN 199 284
FT DOMAIN 290 375
FT DOMAIN 384 625
FT DOMAIN 90 90
FT CARBOHYD 126 126

```



QY	301	PLNNMHTATFAGILIRGSEMPFAGAYOEKYSIHPNYSKRNNDIALMLOKPLTFNDL	360
Db	289	PLSGPRYTATARGILIRGSLMFRGSHOYEKYSIHPNYSKRNNDIALMLOKPLTFNDL	358
QY	361	VKPVCLPMPGMMLOPPEOLCWIISGWATEEKGKTSSEVLNAKVLLETORCNSRYVDNLI	420
Db	359	VKPVCLPMPGMMLOPPEOLCWIISGWATEEKGKTSSEVLNAKVLLETORCNSRYVDNLI	418
QY	421	TPAMICAEFLGNDSCGGDSGGLVTSKNNIMWLIGDTSNCSGAKAVRPVGNWYF	480
Db	419	TPAMICAEFLGNDSCGGDSGGLVTLTKNGIMWLIGDTSNCSGAKALRPVGYDVTVE	478
QY	481	TDWIYROMRAD 491.	
Db	479	TDWIYROMRAD 489	
RESULT	4		
Q9NZAS		PRELIMINARY:	PRT: 423 AA.
AC	Q9NZAS		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)		
DE	TYPE II MEMBRANE SERINE PROTEASE.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_Taxid:9606;		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RT	Smeekens S.S., Lorimer D.D., Wang E., Hou J., Linnevers C.;		
RT	"MT-Sp2, a novel type II membrane serine protease expressed in		
RT	trachea, colon, and small intestine: identification, cloning, and		
RT	chromosomal localization.";		
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF216312; AAF31436.1; .		
KM	Protease.		
SO	SEQUENCE 423 AA; 46397 MW; 90792AF0F8AFE30 CRC64;		
QY	61	SNPVVCTOPKPSGSGVCTKTKKALCTLTGTFVYGAALAGLTKFMGSKCSNGIEC	120
Db	2	SNP--CANPVPD--WRPSESVGIPITIALISLSTIIVYLVILDKYI-----FLC	50
QY	121	DSSGTCINPSMNCDSVSHCPGGEDENRCVRLY--GP-----NFLQVYSQKRSWH	169
Db	51	GQPLHFIRKQKQDDELDPGLGEDHEHCYKSPPEGPAAVRLSKDRSTLOYDSATGMWF	110
QY	170	PVCODDMMENENGRACRDMGY--KNNFSSQ-----GIYDSSSTSMKNTAGAND	220
Db	111	SACFQNFTEALAEYTCRQMGYSKPTFAVEYGPDDIDVYEITENSQELRNRRNSG--	167
QY	221	IYKKLIYHSDAGSSKAVVSLRACIACVNNSSHQSRIVGGESALPCAMPQVSLHYQNVHY	280
Db	168	-----PCLISGSLVSLHCLACKSL---KTPRYVGGEEASVSWPQVSIOTDKOHV	215
QY	281	CGGSITPEWITVTAACHCEKPLNNMHTATFAGILR--OSFMFYGAGYOEKYSI---SHPN	336
Db	216	CGGSITLDPHWVLTAAHCFRKH--TDVFNKMKVRAGSDKLSG---PSLAVAKIITIEFNPM	270
QY	337	YDSKRNNDIALMLOKPLTFNDLVKFPICLPMPGMMLOPPEOLCWIISGWATEEK--GKTS	395
Db	271	Y--EKNDIDIALMLOKPLTFSGTVRPICLPFEDDELTPATPLMTIWMGFTQNGCRKMD	327
QY	396	VLNAKVLLETIRORCNSRYVDNLIPTAMICAEFLGNDSCGGDSGGLVTSKNNIMWL	455
Db	328	ILQLASVOYIDSTRCNADDAIAGEVTEKMKMGKTSPEBGVDTQGGDSGGLVTSKNNIMWL	455

Query	Best Local Similarity	Matches	Conservative	Score	DB	Length	Indels	Gaps
133 CDGVSHCPGEGEDENFCVRLY--GP-----	24.94	39.18	57	676.5	DB 4	437	49	13
77 CDGEIDCPGLGEDEEHCHVKSPEEGPAVAARLSKDRSTLOVLDSATGCMFMSACFDNFTAL	24.94	39.18	57	676.5	DB 4	437	49	13
182 RAACGDMY--KNNYSSQ-----GVDDSGSFMKLTMSAGNVDIYKKLYSDACS	24.94	39.18	57	676.5	DB 4	437	49	13
137 ETACQKQKISSPFFRAIEIGDDDLVDVETENSQELMRNSG-----PCL	24.94	39.18	57	676.5	DB 4	437	49	13
233 SKAAYSLERICAGVNLNRSROSRIYGESEALPGAPMCOVSLHNOVHYVCGSITPENIV	24.94	39.18	57	676.5	DB 4	437	49	13
185 SSSLYSLHCLACGKSL---KTPRYVGGESADVSMPQVSIQYDKQHGCGSIIIDPHVYL	24.94	39.18	57	676.5	DB 4	437	49	13
293 TAACVEKPLNNPWHVTFAGILR--QSMWFGAGYQVEKYV---SHPNYDSKTKNDIAL	24.94	39.18	57	676.5	DB 4	437	49	13
242 TAAHCFRRH--TDVFNWKKRAGSGDKLGSF---PSLAVKAIIEENPNY---PKNDIAL	24.94	39.18	57	676.5	DB 4	437	49	13
349 MKLOKPLFENDLVKVCVCPNPGMMLQAPOLQISMGAGTEEK--GKTSVLNAKULLIET	24.94	39.18	57	676.5	DB 4	437	49	13
234 MKLOPLFLFSGVRICTLPFDEELTPATPLMITISMGFTKQKQMSIDILQASQVIVDS	24.94	39.18	57	676.5	DB 4	437	49	13
408 QRCNSRYVYDNLITPAMICAGFLQNVDSQSDSGPLVTSKNNIMLIGDTSWGGCAK	24.94	39.18	57	676.5	DB 4	437	49	13
354 TRCNADDAVQGEVTEKMKACAGIPGEGVDTQCGDSSGPLMYGSDQ--WHYVGIYSMAGCGG	24.94	39.18	57	676.5	DB 4	437	49	13
468 AYRPGVYGNVYFTDWYTRQNRAD	24.94	39.18	57	676.5	DB 4	437	49	13
413 PSTPGVYTKVSAYLNMVYVWKA	24.94	39.18	57	676.5	DB 4	437	49	13



```

OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=JEJUNUM;
RA Tsuzuki S.;
RT "A membrane bound serine protease expressed in rat small intestine.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB037898; BAB03502.1;
KW Protease.
SQ SEQUENCE 855 AA: 94955 MW: 358067E6FC6F03D CRC64:

Query Match 20.8%; Score 564.5; DB 11; Length 855;
Best Local Similarity 33.0%; Pred. No. 1e-42;
Matches 128; Conservative 62; Mismatches 131; Indels 67; Gaps 11.

OY 110 GSKCSNGSIECDSSGTCINPMKCDGVSHCPGGEENRCVRLYGPNFIQVYSSQKSMH 169
DB 522 GCSCPAGSFKC-SNGCLQLPQSQCGKCKDCGSDPASCDDNVNAVSCRTKYTRCO----N 576
OY 170 PYCGDDMN-ENTGRACRDMGYKNNFYSSQGLVDSGSTSEPKLNTSAGNDIYKKLYHS 228
DB 577 GCLNKGNEPCGCKKDCSGSDEKNC-----DCGLSFTRK----- 611
OY 229 DACSSKAVVSLCINACGVNLSSROSRIYGGESALPGAPMOWSLH-YONVHVCGSIT 287
DB 612 -----QARVYGGTNADEGEMPMOVSLLHALGCHLGSALIS 647
OY 288 PEMIVTAACHVEK---PLNNPMHTAFAGILROSEMFYGAQYQ--VEKVISHPNYDSK 340
DB 648 PMIVSAALHCPQDEFETFKKSDHTMTWTAFLGLLDQS-KRASGVGEHKLRITHTPSFNF 706
OY 341 TKNNIDALMKLQPLTEFNDLVKPVCLPNFGMLQPEQLCWISGWTGTEKTKTSEYLNAA 400
DB 707 TEDYDIALLELEKPAEYSTVYRPICLPDHTHFEPAKAIWTVGWGHTKEGTCGALILIKG 766
OY 401 KVLLETQCNRSRYVDNLTIPAMICAGELQCNVDSGCCDSGELPT-SKNINIMWLTGDT 459
DB 767 EIRVINQTTCEE--LLPQGITPRMCCVGLSGGVSCDSGSGPLSSVERDGRIFQAGVY 824
OY 460 SMGSGCAKAYRPGVYGNVAVFTDVIYRQ 487
DB 825 SMGEGCAQRNKPQVYTRIREVARDWIKEO 852

RESULT 9
OY7506 PRELIMINARY: PRT: 643 AA.
AC O97506;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE KALLIKREIN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLASMA;
RA Takehashi T., Kimura A., Okimura H., Hamabata T.;
RT "Porcine liver plasma kallikrein.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR HSR: AB022425; BAA37147.1;
DR HSR: P00766; ICHG.
DR MEROPS: S01.212; -.
DR INTERPRO: IPR001177; -.
DR INTERPRO: IPR001254; -.
DR INTERPRO: IPR001314; -.
DR INTERPRO: IPR003014; -.
DR PFAM: PF00024; PAN: 4.
DR PFAM: PF00089; trypsin; 1.
DR PRINTS: PR00005; APPLEDOMAIN.

```

```

DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
DR PROSITE; PS00495; APPLE; 4.
SQ SEQUENCE 643 AA; 7227 MW; AFF2923E3C3CB80A CMC64;

Query Match 20.7% Score 563.5; DB 6; Length 643;
Best Local Similarity 29.9%; Pred. No. 8,86-43;
Matches 154; Conservative 63; Mismatches 185; Indels 113; Gaps 19;

QY 52 YAPRYLTQASNPVYCTQPSGTYCTSKTKALCTLTGLTGVLAALACLKFKMG 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 YATQAFNNAEYRNKCNLTKHSPGFTPTSIK-----VLAVESG-----FSLK 200
QY 112 KCSNSGICDSDS-----GTCINPSMCCGVCCHCPGDEBNCYRLYGNFILQ 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 PCADSEICCHNDIFQHLAFSDVDVARVAPAFV-----C-----PTICYHPNCLFF 248
QY 160 VY-----SSOR-----KSWH-----PVCODDNNENYGRAACRDM---GYKNPFYSQ 198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 TFFYNAMAKIESRNVCFLEKTSHGSPFTPEQNAISGYSLTTCQTLPEPCHSKITSEY 308
QY 199 GIVDDSGSTSEF-----KLNTSAGNVDIYKKLYHSDAC--SSKAVSLR----- 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 DFESELVAVTFQGANLQGEETCTKTRIQGFYSLHPEDCREKCKSLRLSDGSPTRKI 368
QY 241 -----C-----IACGVNLNLSRROSRIYVGESALPGAMPQVY-----LHYON 277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 THGMRASSGYSLRLCRSDHSHCAATKAN---TRIVGTDSFLGEMPMQVYSLOAKLRAON 424
QY 278 VHYGGSITTEPMYITAAHCHVERKLNPNMWHMTAFAGILLROSMFPGAGV--QYEVKISHPN 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 -HLCGGSTIGHQWVLTAAHCFD--GLSLDPIMWITVGLINLSITETPEPSQVKEITIHON 482
QY 337 YDSKKNNDIALMKLQRFLETFENDLYKPVCLPFGMMLPEQCLWISGWTGEKSTSEY 396
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 KYLESQHDIALKLETFPLNTDPRKPICLPBRDNTNVTTCWTKGKFTKEKEIOMI 542
QY 397 LNAAKVLLIEQRCNSRYVDNLITPAMICAGFLOGNVSDCGDSGGPLVTSKNNIMWLI 456
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 543 LQKVIAPLVSNCEQCKY--RDHKISKQMICAGYKKGKADCKGSGGGLVCKYNGIMHLY 601
QY 457 GDTSMGSCAKAIRPGVYGNVAFEDWITRQMRAD 491
Db 602 GTTSMGEGCARREDPGVYTKVLEYMDWILEKTQDD 636

RESULT 10
ID Q9Y5Y6 PRELIMINARY; PRT; 855 AA.
Q9Y5Y6
AC Q9Y5Y6;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE MATRILYASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
CX NCBI_TaxId:9606;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE:99303581; PubMed:1037424;
RA Lin.C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
RT "Molecular cloning of cDNA for matrilysin, a matrix-degrading serine
RL J. Biol. Chem. 274:18231-18236(1999).
RN [2]
RN SEQUENCE FROM N.A.
RA Lin.C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
RN Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RP [3]
RP SEQUENCE FROM N.A.
```



Db	922	AKCEKRGDEIYCGDS-IGCIGTKHICDDIIDCPYGDDEHNCRLRLSERNGDYGTGLVEYLR	980
Qy	163	SQRKSMHPVCODDMNENY-RAACRDMYKNNFYSSQGIYDSSGTSFMK-----LNT	214
Db	981	IGOROMPACVKNMRAVSPSAVCISILGY-----SAVNATSVLTQTHRPPLAT	1029
Qy	215	SAGVNDIYK-----KLHSDACSSKA-----VSLRC--INAG-VNLNNSQGS-RITYG	259
Db	1030	VNVSITDMKKYAKRKSTLMQEFNCKRTEDYPADLTCSNYECGRYKGRNHPSSRIIG	1089
Qy	260	ESALPGAMPQVSL--HYQNVHVCGSITTEPMIYTAHCY-EKRLNPMHTTAAGILR	316
Db	1090	TQSPGMMPLAIIICGPEKIFTCACVLLISDQVLTASHCAGNYSVIDEDMTIGLGYTR	1149
Qy	317	Q-SFEMFGAGYQYQEKYISHPNYD-SKTKNNIDIALMKRLFLFNLYKPVCLPNPMM-L	373
Db	1150	RNSFTYSIGQYKKKAVIAPHQYIMALAHNDIALDOLATRAVAFHEHLRPVCLPPPSVNL	1209
Qy	374	QPEOLCWSMGATEEKGKTS--VLNAKYLLIETORCNSRYVYDNL-ITPAMICAGF	429
Db	1210	HPGLTCLVIMGCRKEDPRSTVEYIVNEQVPIITRNQCD--WLDNLTYSEGWVCAGF	1267
Qy	430	LOGNVDSQCGDSGPLY--TSKNNITMWLTIGTSMGSCAKAIRPGYGNVAFETDMLR	486
Db	1268	DDGKGKDACQDSDSGPLCPYRPGKRNRMFVGIVSWGIMCAHPRLTGGVYANVQYVPWIOE	1327
Qy	487	QM 488	
Db	1328	QI 1329	

RESULT	12	
09Y495		
ID	09Y495	PRELIMINARY;
AC	Q3Y495;	PRT; 571 AA.
DT	01-NOV-1999 (TREMBLrel, 12, Created)	
DT	01-NOV-1999 (TREMBLrel, 12, Last sequence update)	
DT	01-OCT-2000 (TREMBLrel, 15, Last annotation update)	
DE	PLATELET FACTOR XI.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxId:9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MDLRef:98256306; Pubmed:9593722;	
RA	Hsu T.C., Shore S.K., Seshsma T., Bagasra O., Walsh P.N.;	
RT	"Molecular cloning of platelet factor XI, an alternative splicing	
RT	product of the plasma factor XI gene.";	
RL	J. Biol. Chem. 273:1787-1793(1998).	
DR	EMBL; AF045649; AAC24506.1; -.	
DR	HSSP; PF02031; IAAO.	
DR	INTERPRO; IPR00177; -.	
DR	INTERPRO; IPR001254; -.	
DR	INTERPRO; IPR001314; -.	
DR	INTERPRO; IPR003014; -.	
DR	PFAM; PF00024; PAN; 3.	
DR	PFAM; PF00089; trypsin; 1.	
DR	PRINTS; PR00005; APLEDDOMAIN.	
DR	PRINTS; PR00722; CHYMOTRYPSIN.	
DR	PROSITE; PS00134; TRYPSIN_HTS; UNKNOWN_1.	
DR	PROSITE; PS00135; TRYPSIN_SRR; 1.	
DR	PROSITE; PS00495; APPLE; 3.	
QO	SEQUENCE 571 AA; 63840 MW; FB9D65D72151755E CRC64;	

	Query Match	19.6%	Score 533	DB 4	Length 571	
	Best Local Similarity	30.5%	Pred. No.	4.5e-40		
	Matches 122:	Conservative	63	Mismatches	129	Indels 86
	Gaps	11				
OY	169 HPVC-----QDDNHNENGRACRGMGKNNNYSSQGIVDSDGSFMSFKLNTSAGVNDIY	222				
	:       :       :       :					

[illegible]

DE DISTAL INTESTINAL SERINE PROTEASE.  
GN DISP.  
OS *Mus musculus* (Mouse).  
OC *Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;*  
OC *Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.*  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Shaw-Smith C.J., Coffey A.J., Laversha M., Freeman T.C., Bentley D.R.  
RA Walters J.R.;  
RT "Characterization of a novel murine intestinal serine protease,  
RT DISP.":  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ243866; CAB56465.1; -.  
DR HSSP; P20231; IAAO.  
DR INTERPRO; IPR001254; -.  
DR INTERPRO; IPR001314; -.  
DR PFAM; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PROSITE; PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
KW Protease.  
SQ SEQUENCE 310 AA; 33701 MW; F82BEC7F6D25303F CRC64;

```

Query Match      19.8%; Score 529; Db 11; Length 310;
Best Local Similarity 41.7%; Pred. No. 4.8e-40;
Matches 109; Conservative 42; Mismatches 91; Indels 14; Gaps

OY 244 CGVNLNSRROSRIYGGESALPGANPMVOYSLHY-QNYAVCGSGSITTEBWIYTAHAHCYEKPL 302
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 28 CG---HSRDAGKIVGGDDALEGOMPVOVSLMTIEDGICGSLIHEWVYVLAACHRRSL 84
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

OY 303 NNPWHTAFAGILROSEMFYAGAGYOEKVISHPNY-DSKKNNDIALMKLQKLPTEFNDLV 361
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 85 NPSTYHKVAGGLTSLLEPRTSLVAVNIIHPHTLYLMAADSSDIALVQDPLPRLRSQFT 144
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

OY 362 KPVCLPFGMMLDPEQLCMISGMAATEEKKTSIEVLAAKVLIIENFQRCNSKRL----- 414
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

